Nonadaptive fault-tolerant verification of quantum supremacy with noise

Nonadaptive fault-tolerant verification of quantum supremacy with noise

Theodoros Kapourniotis and Animesh Datta

Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom

June 27, 2019

Quantum samplers are believed capable of sampling efficiently from distributions that are classically hard to sample from. We consider a sampler inspired by the classical Ising model. It is nonadaptive and therefore experimentally amenable. Under a plausible conjecture, classical sampling upto additive errors from this model is known to be hard. We present a trap-based verification scheme for quantum supremacy that only requires the verifier to prepare single-qubit states. The verification is done on the same model as the original sampler, a square lattice, with only a constant overhead. We next revamp our verification scheme in two distinct ways using fault tolerance that preserves the nonadaptivity. The first has a lower overhead based on error correction with the same threshold as universal quantum computation. The second has a higher overhead but an improved threshold (1.97%) based on error detection. We show that classically sampling upto additive errors is likely hard in both these schemes. Our results are applicable to other sampling problems such as the Instantaneous Quantum Polynomial-time (IQP) computation model. They should also assist near-term attempts at experimentally demonstrating quantum supremacy and guide long-term ones.

1 Introduction

Considerable experimental efforts are being directed towards the realisation of quantum information processing technologies, with the eventual aim of constructing universal quantum computers and simulators [35, 37, 48, 49]. One of the motivations for this exercise is their expected ability to simulate physical systems that are believed to be intractable classically [18, 27]. An example of such a system is a lattice of interacting spins in the presence of a magnetic field, represented by the classical Ising model [44], which is a workhorse in condensed matter physics and statistical physics [8, 51, 66].

Computing the partition function of the Ising model in an external magnetic field is, however, #P-hard even in two spatial dimensions with multiplicative approximations [34, 39]. #P-hard problems are not expected to be solvable efficiently on a universal quantum computer. *Sampling* up to multiplicative [9, 13, 25, 70] and additive [1, 7, 12, 36, 56, 58] errors from certain distributions, such as from the partition function of the Ising model at imaginary temperatures, is possible using devices that do not require the full complement of DiVincenzo's criteria [23] for their implementation. Their scalable implementation is thus anticipated to be more achievable than a universal quantum computer's, providing tangible demonstrations of quantum supremacy sooner [9, 15, 32, 50, 68]. This expectation is purchased at the price of sacrificing the full power of universal quantum computers, promising only to efficiently sample from certain distributions instead. This is the remit of 'quantum supremacy' [62] but the relative experimental ease introduces new theoretical challenges [41].

Given the significance of quantum supremacy in wider quantum information science, demonstrating it experimentally is vital. In the real world however, this faces two crucial challenges [2, 41] that arise from experimental imperfections and noise respectively. The first is verifying that the output distribution is correct, or at

Theodoros Kapourniotis: T.Kapourniotis@warwick.ac.uk
Animesh Datta: animesh.datta@warwick.ac.uk

least close to correct. Since any real world experiment will be imperfect, establishing correctness is important, particularly so since sampling, unlike for instance integer factoring, is not in NP whose correctness can be checked efficiently. This calls for a verification scheme with minimal trust assumptions, that indicates whether or not the output distribution is sufficiently close in total variation distance to the ideal one, and consequently whether or not an experiment has successfully demonstrated quantum supremacy. The hurdle is to ensure that the verifiable supremacy experiment no more demanding than the original non-verifiable experiment. This requires a redesign of verification schemes for universal quantum computing [4, 16, 29].

The second challenge arises because all experiments are noisy. Arguing for experimental quantum supremacy is incumbent on the sampling task being classically hard to simulate even in the presence of noise. As excessive noise can render a hard probability distribution easy to simulate, it is an important challenge to determine to what extent a sampling task remain hard to simulate classically, even in the presence of noise [41]. The hurdle therefore is to retain hardness in the presence of noise. This is what fault-tolerance provides. Of course, this is only worthwhile for experiments if the fault-tolerance threshold for verifiable quantum supremacy is strictly easier to achieve than that of universal quantum computation. Furthermore, the overheads needed for fault-tolerant verifiable quantum supremacy must be less demanding than that for universal quantum computation. Fault-tolerant quantum supremacy is thus a compelling milestone on the way to a fault-tolerant universal quantum computer.

Among the numerous quantum supremacy candidates [41] are IQP [13] and those allowing sampling from the distribution of partition functions of the classical Ising model at imaginary temperature - the Ising sampler [36]. IQP and the Ising sampler exhibit in line with other models [1, 9, 25, 34], under plausible conjectures, a highly unexpected collapse of the polynomial hierarchy to the third level occurs if a classical sampler can sample the partition function distribution upto additive errors. The Ising Sampler is a constant-depth version of IQP, with the additional favourable properties of translational-invariance and single-instance. Single-instance means that the hardness results hold for a single fixed instance of the problem, relieving the burden of creating random instances such as Boson Sampling [1] and Random Sampling [9]. This simplifies both theoretical analysis and experimental implementation. Although the partition function at imaginary temperatures may appear unphysical, it has deep connections to quantum complexity theory [39] as well as quantum statistical and condensed matter physics via analytic continuations [51]. It is the combination of relative ease in theoretical analysis and experimental implementations allied with its strong connections to physics that motivates our choice of the Ising Sampler as the subject of this paper.

In this paper, we provide a fault-tolerant scheme for the verification of quantum supremacy in the Ising sampler. We achieve this by amalgamating trap-based quantum verification techniques [29] with recent results on demonstrating quantum supremacy by emulating fault tolerance via post-selection [32]. In response to the first challenge above, we present a nonadaptive verification scheme with exponentially low probability of failure and only linear complexity for the Ising sampler (prover). Our scheme applies to any untrusted prover with entangling and measuring capabilities, limited only by the laws of quantum mechanics, and requires the verifier to prepare random, single-qubit states with bounded local noise. We first present a verification scheme that should aid demonstrating quantum supremacy with few qubits (Theorem 1). In response to the second challenge above, we prove fault-tolerant versions (Theorem 2), one of which uses the idea of emulated fault tolerance by post-selection [32]. This 'free' post-selection enables us to provide a fault-tolerant verification scheme with improved thresholds over universal quantum computing thresholds. An important property of our verification is that it itself is within the instantaneous model of quatum computing and therefore can be implemented in the same device as the Ising sampler with small modifications. Moreover, we prove quantum supremacy of this modified model (Theorem 3).

1.1 Comparison to prior work and structure

We go beyond Ref. [29] in three ways, namely (i) providing a new definition of verifiability over

many i.i.d. repetitions of the protocol, based on the total variation distance between the output distribution and the correct one; (ii) using the Raussendorf-Harrington-Goyal (RHG) strategy for fault tolerance instead of using it for probability amplification; and (iii) developing a simpler construction for verification and computation on separate planar graphs.

We go beyond Ref. [32] by (i) solving one of its open problems – proving quantum supremacy of improved threshold fault-tolerant model up to *additive* errors (Theorem 3) and (ii) verifying quantum supremacy while maintaining its improved thresholds.

Our work is a significant improvement over the verification method of Ref. [36] on two counts. Firstly, our schemes require a linear overhead in the number of qubits as opposed to a quadratic one in Ref [36]. Secondly, our schemes (fault-tolerant ones) scale in the face of constant local noise while that of Ref [36] requires local noise polynomially small in the total number of qubits. Our verification schemes apply to any nonadaptive sampler based on cluster states, however we will use the Ising sampler [36] as a particular example, thus keeping the benefit of the single instance property and experimental feasibility of this model.

Our work is structured as follows. Section 2 defines verifiability based on the total variation distance of the output distribution. Section 3 provides an overview of the Ising sampler placed in an a cryptographic setting of a prover and a verifier. Section 4 contains the first of our main results on the verification of the Ising sampler's output. We present a non-fault-tolerant verification scheme (Theorem 1). Since it requires decreasing noise in preparation, entanglement and measurement with increasing system size, this is only viable in small-sized experimental demonstrations of quantum supremacy. In Section 5 we present two fault-tolerant versions of the verification scheme (Theorem 2) which are scalable when noise is below certain thresholds. Section 6 provides a result (Theorem 3) on the quantum supremacy of the output distribution of the noisy Ising sampler conditioned on syndrome measurements accepting, which is a generalisation of Ref. [32] for additive errors.

2 Verifiability

We begin with our definition of a verification protocol.

Definition 1 (Verification protocol)**.** *A verification protocol involves two parties - a trusted verifier and an untrusted prover who share a quantum and classical channel. The protocol takes as input a description of a computation and outputs a string and a bit. The bit determines if the string is accepted or rejected.*

Establishing verifiability of a protocol consists of proving completeness and soundness. A protocol is complete if, for an honest prover, the verifier outputs the correct result and accepts. A protocol is sound if, for any deviation of the prover, the probability that the verifier outputs an incorrect result and accepts is low. This deviation captures both a malevolent prover who tries to cheat and uncontrollable errors in the prover's device.

Note that the above notion of verifiability relies on an output string being correct while sampling relies on distributions being close. We are therefore interested in the total variation distance between the experimental output distribution and the exact one [29]. We are furthermore interested in arguing for quantum supremacy based on the total variation distance between distributions. This requires us to go from a joint distribution of a string and a bit to a probability distribution over strings conditioned on a bit. To meet these demands we introduce the idea of a verification scheme, that uses a protocol as a black box and can call it repeatedly. We also assume that the repetitions of the protocols are independent and identically distributed (i.i.d.). However, there is no assumption on the behaviour of the system within the protocol, which means that an adversarial prover can cheat by correlating systems within the protocol.

Definition 2 (Verification scheme)**.** *A verification scheme takes as input a verification protocol, $M \in \mathbb{N}, l \in [0,1]$ and outputs a string and a bit. The bit determines if the string is accepted or rejected.*

A verification scheme works as follows. After running M i.i.d repetitions of a verification protocol it outputs one of the M output strings at random and accepts if at least a fraction l of the protocols accept and rejects otherwise.

Let $q^{\text{nsy}}(\boldsymbol{x})$ be the experimental and $q^{\text{exc}}(\boldsymbol{x})$ be the exact distribution of the output $\boldsymbol{x}$ of a sampler. We are interested in the quantity

$$\text{var} \equiv \frac{1}{2}\sum_{\boldsymbol{x}} |q^{\text{exc}}(\boldsymbol{x}) - q^{\text{nsy}}(\boldsymbol{x})|, \tag{1}$$

where the sum is over all binary strings $\boldsymbol{x}$ of size N.

The following definition captures the notions of completeness and soundness at the level of a scheme for sampling problems.

Definition 3 (Verifiability of a scheme)**.** *A scheme is verifiable if its output is*

- $(\delta', \delta)-$*complete: For an honest prover having only bounded noise, the scheme accepts at least with probability* δ'*, and*

$$\text{var} \leq 1 - \delta \tag{2}$$

for the output string.

- $(\varepsilon', \varepsilon)-$*sound: For any, including adversarial, prover if the scheme accepts then*

$$\text{var} \leq \varepsilon \tag{3}$$

with confidence ε'*.*

We then consider the verifiability of a scheme for a sampler which has a designated output register we call the post-selection register. We consider probabilities $q^{\text{nsy}}(\boldsymbol{x}|y=0)$, where y is the value of the post-selection register, for the experimental and $q^{\text{exc}}(\boldsymbol{x}|y=0)$ for the exact distribution of a sampler, conditioned on y being zero. We are interested in the quantity

$$\text{var}^{\text{Post}} \equiv \frac{1}{2}\sum_{\boldsymbol{x}} |q^{\text{exc}}(\boldsymbol{x}|y=0) - q^{\text{nsy}}(\boldsymbol{x}|y=0)|. \tag{4}$$

Again, the sums are over all binary strings $\boldsymbol{x}$ of size N. We adapt our definition to conditional probabilities as follows.

Definition 4 (Verifiability of a scheme for post-selected distribution)**.** *A scheme is verifiable conditioned on the post-selection register being zero, if its output is*

- $(\delta', \delta)-$*complete: For an honest prover having only bounded noise, the scheme accepts at least with probability* δ'*, and*

$$\text{var}^{\text{Post}} \leq 1 - \delta \tag{5}$$

for the the output string.

- $(\varepsilon', \varepsilon)-$*sound: For any, including adversarial, prover if the scheme accepts, then*

$$\text{var}^{\text{Post}} \leq \varepsilon \tag{6}$$

with confidence ε'*.*

3 Quantum sampling in the verifier-prover setting

In the verifier-prover setting, the verifier can prepare bounded-error, single-qubit states and the prover implements the rest of the computation including the measurements, and returns the output samples to the verifier. The role of the verifier is to ascertain if the prover is acting honestly and executing the correct operation. The prover is, in general, malicious, trying to pass any tests designed by the verifier while deviating from the correct implementation at the same time. This malice may be intentional if the prover is trying to convince the verifier of its quantum power when it has none, or incidental if the prover possesses an imperfect quantum device prone to noise and errors. We assume that the prover's deviations are governed by quantum mechanics.

We begin by adapting the Ising spin model to a blind verifier-prover cryptographic setting [17]. Blindness, which ensures that the prover remains ignorant of the actual computation, is a necessary ingredient in our verification scheme. Our Ising spin model consists of qubits in state $|+\rangle$ subject to nearest neighbour controlled π-phase rotations, denoted by cZ. All the qubits are measured simultaneously in a basis in the xy-plane of the Bloch sphere. The measurement outcome of classical bits is the output sample. This model corresponds to the well-studied measurement-based quantum computing (MBQC) model [64] without the adaptive measurements. This last restriction makes the depth of the computation constant on the size of the input: one round of preparation, three rounds of entangling because the maximum degree of the graph is three and one round of non-adaptive measurements. This relaxes DiVincenzo's criteria of long decoherence times and makes this model non-universal for quantum computing.

In the particular Ising sampler presented in [36] the structure of the graph state is fixed (Fig. (1)), but its size scales with the width m and depth n. The measurement angles are also fixed to specific

values from the set $\{-\pi/4, -\pi/8, 0, \pi/8, \pi/4\}$. This choice of graph, which we call the 'extended' brickwork state, and a fixed angle for each physical qubit has the following benefit: Each possible combination of measurement outcomes 'chooses' a different angle for each qubit of the original brickwork state from the set $\{k\pi/4\}$, $k = \{0, \ldots, 7\}$. This effectively makes a single instance of the model a random quantum circuit generator, a property exploited to prove its hardness.

The correspondence to an Ising model comes from the locality of spin interactions and decomposing each MBQC measurement into a unitary rotation around the z-axis corresponding to an external magnetic field, followed by a Pauli X measurement. The quantum state just before the Pauli X measurement is given by the unitary evolution due to the Hamiltonian

$$\mathcal{H} = -\sum_{\langle i,j \rangle} J Z_i Z_j + \sum_i B_i Z_i \tag{7}$$

where J is the interaction term, B_i the local field strength and Z_i the Pauli Z operator on qubit i.

The probability $q^{\text{exc}}(\boldsymbol{x})$ of measuring a bit string $\boldsymbol{x}$ corresponds to the partition function $\mathcal{Z}_{\boldsymbol{x}}$ of the Ising model with Hamiltonian $\mathcal{H}' \equiv \mathcal{H} + \frac{\pi}{2}\sum_i x_i Z_i$ and is given by

$$q^{\text{exc}}(\boldsymbol{x}) = \frac{|\text{Tr}(e^{-i(\mathcal{H}+\frac{\pi}{2}\sum_i x_i Z_i)})|^2}{2^{2N}} \equiv \frac{|\mathcal{Z}_{\boldsymbol{x}}|^2}{2^{2N}}, \tag{8}$$

where $N = mn$. The second term in $\mathcal{H}'$ comes from the measurement outcomes of the Pauli X measurements, and the partition function is evaluated at an imaginary temperature $\beta = 1/k_B T = i$.

Testing the honesty of the prover, in our case the Ising sampler, requires the 'blind' injection of certain 'trap' qubits. To keep the identity of these trap qubits from the prover, the verifier applies some encoding on the original translationally-invariant Ising spin model, making the model translationally variant. Now both the participating qubits and the measurement angles on the graph state have a randomly chosen extra rotation according to the scheme described next.

Specifically, each qubit i is individually prepared by the verifier in the state $|+_{\theta_i}\rangle$, where θ_i is chosen uniformly at random from the set $A = \{0, \frac{\pi}{8}, \frac{2\pi}{8}, \ldots, \frac{15\pi}{8}\}$. Instead of the prover measuring in fixed predetermined angles, as in the original Ising sampler, the verifier sends encrypted angles to the prover: $\delta_i = \theta_i + (-1)^{r'_i}\phi_i + r_i\pi$ for $r_i, r'_i \in_R \{0, 1\}$, where $\in_R$ stands for a uniform random selection. Rotations by θ_i on the qubit and on the angles mutually cancel and the classical information that the prover receives (containing the actual measurement angles ϕ_i) is classically one time padded by θ_i. The bits r_i, r'_i provide some extra randomness to restrict the information the prover gets from the quantum state and can be corrected by classical post-processing of the sample. Our difference from Ref. [17] lies in the number of angles used in the set A, and comes from the fact that we use a different decomposition of the computation. We conjecture that this can be improved upon (See Sec. 7).

4 Non-fault-tolerant Verification of Ising Sampler

The output of a quantum sampler must be classical for it to be comparable to that of a classical sampler, a prerequisite for demonstrating quantum supremacy. This allows us to simplify trap-based verification strategies for universal quantum computation [29, 46, 47] to having disjointed computational resource and trap states - an idea also used in Ref. [47] and in circuit-based verification [16]. This permits an exponentially small error in our estimation of the fidelity of the output using a square lattice. Finally, a trap-based technique instead of fidelity-witness based certification ones [36, 40], similar also to [5, 20], enables us to reduce the resource complexity of the verification protocol from quadratic to linear. Other certification methods that require linear resources exist, by trusting the measuring devices [26, 42, 54, 60] instead of preparation. Linear resource complexity is minimal in this scenario because the verifier needs to receive at least one copy of the resource state to perform the computation.

Our verification protocol relies on judiciously selecting the measurement angles and placing *dummy* qubits prepared in the state $|0\rangle$, which, together with xy-plane measurements, allows us to *carve* different types of graphs from a square lattice graph, as shown in Fig. (3). Placing one dummy qubit between any two other qubits prevents the prover's entangling operators to have any entangling effect between the participating qubits, so the prover can apply exactly the same

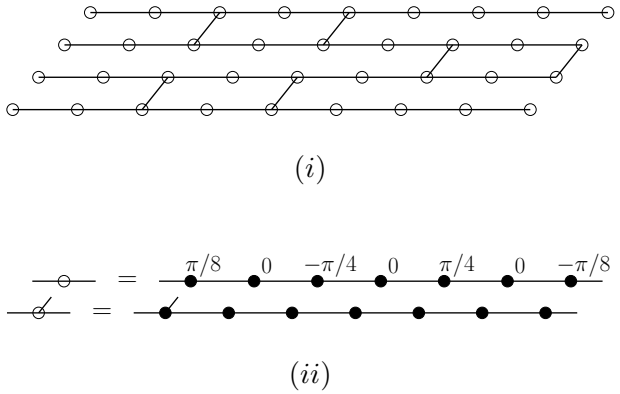

Figure 1: The original brickwork state (i) is a universal resource for MBQC under xy-plane measurements, where white vertices represent qubits and the edges represent cZ operations. The 'extended' brickwork state (ii) is used in the original Ising sampler [36], where each white vertex is replaced by 7 physical qubits (black vertices). The measurement angle for each qubit is fixed to the value written above each vertex. There is no adaptation of the angles based on previous measurement outcomes as in universal MBQC.

operations that produce the square lattice but create a different graph state. The graphs carved out are the 'extended' brickwork state (Fig. (1)) and two other graphs containing special 'trap' qubits in the state $|+\rangle$. The extended brickwork state is used to run the Ising sampler. The traps in the trap graphs are measured in the same basis as prepared, yielding a deterministic check on the prover. Two different types of the trap graphs are needed to enable placing a trap at any position in the graph with equal probability. The order of the graphs is chosen at random and the whole protocol implemented blindly to thwart the prover from distinguishing trap and target computation qubits.

A sketch of our Protocol 1 appears in Fig. (2), and the details in Sec. 4.1. The protocol has constant time complexity of the quantum operations and needs $O(N)$ qubits, where N is the number of the qubits of the Ising sampler.

Noise considered in all our protocols for an honest prover is local, unital and bounded. It applies after every elementary operation (preparation, entangling and measurement) j and is expressed as a CPTP superoperator:

$$\mathcal{N}_j = (1 - \epsilon_{V,P})\mathcal{I} + \mathcal{E}_j \tag{9}$$

where $\epsilon_{V,P} = \epsilon_V$ and $||\mathcal{E}_j||_\diamond = \epsilon_V$ for the noise of the verifier (preparation noise) and $\epsilon_{V,P} = \epsilon_P$

1. Verifier selects a random ordering of $2\kappa+1$ graphs, one for target computation and κ from each type of trap graphs.

2. Verifier prepares, one by one, the qubits needed for the blind implementation of the $2\kappa + 1$ cluster states and sends them to the prover.

3. Verifier sends the encrypted measurement angles to the prover.

4. Prover entangles all received qubits in the $2\kappa + 1$ cluster states.

5. Prover measures all qubits simultaneously in the instructed angles and returns the results.

6. Verifier decrypts the outputs and accepts if all trap results are correct, otherwise rejects.

Figure 2: Nonadaptive verification protocol

and $||\mathcal{E}_j||_\diamond = \epsilon_P$ for the noise of the honest prover (entangling and measurement noise).

Theorem 1 (Non-fault tolerance verification scheme)**.** *There exists a verification scheme with Protocol 1,* $M = \log(1/\beta)/(2\kappa^2 N^2(\epsilon_V + \epsilon_P)^2)$, $l = (1-\kappa N(2\epsilon_V+4\epsilon_P))$ *that according to Def. (3) is*

$$\left(1-\beta, 1-\sqrt{N(\epsilon_V+3\epsilon_P)}\right) - \textit{complete}$$

and

$$\left(1-\beta, \sqrt{\kappa N(3\epsilon_V+5\epsilon_P)+\Delta_\kappa}\right) - \textit{sound},$$

where $\Delta_\kappa = \kappa!(\kappa+1)!/(2\kappa+1)!$.

In the above, ϵ_V and ϵ_P are fixed by the experimental capability, while completeness and soundness parameters are set by the conjectures invoked to argue for quantum supremacy, as obtained in Eqn. (18).

A proof sketch appears in Section 4.3 and a full proof in Appendix B.

Using our verifiable quantum sampler to demonstrate quantum supremacy is underwritten by results which show that approximating the Ising sampler upto constant total variation distance is hard classically, subject to an average case hardness and an anti-concentration conjecture, presented in Section 6, similarly to the original model [36].

Both $\kappa N \epsilon_V$ and $\kappa N \epsilon_P$ must be constant for the total variation distances to be constant, plus exponentially decaying in κ term Δ_κ in soundness, in Theorem 1. To achieve this we require local errors ϵ_V and ϵ_P to decrease linearly with the number of qubits and κ. This is only realistic in quantum supremacy experiments involving a few qubits.

To overcome this restriction, we consider fault-tolerant versions of our verification protocol in Section 5.

4.1 Protocol

The following is a full description of the non-fault tolerant verification protocol:

Protocol 1:

1. Verifier selects a random ordering of $2\kappa + 1$ graphs, one for computation and 2κ for testing, as in Fig. (3). This fixes the position of computational basis qubits called the *dummy* qubits (see Appendix A) and the measurement angles $\{\phi_i\}_{i=1}^{N}$, where $N = m \times n$ is the total number of qubits, so that

   (a) in the target computation graph we carve from the square lattice a universal resource state, the 'extended' brickwork state of Figure (1) and fix the rest of the measurement angles according to the Ising sampler model;

   (b) in the trap computation graphs the dummy qubits are used to isolate the traps, which are placed at fixed positions. For half of the graphs in positions with odd parity that correspond to non-dummy qubits in the computational graph and in the other half in positions with even parity that correspond to non-dummies in the computational graph. The traps are measured with angles $\phi = 0$ so that the measurement is deterministic. Crucially, the trap graphs do not contain any 'bridge' operations so there is no need for adaptive corrections.

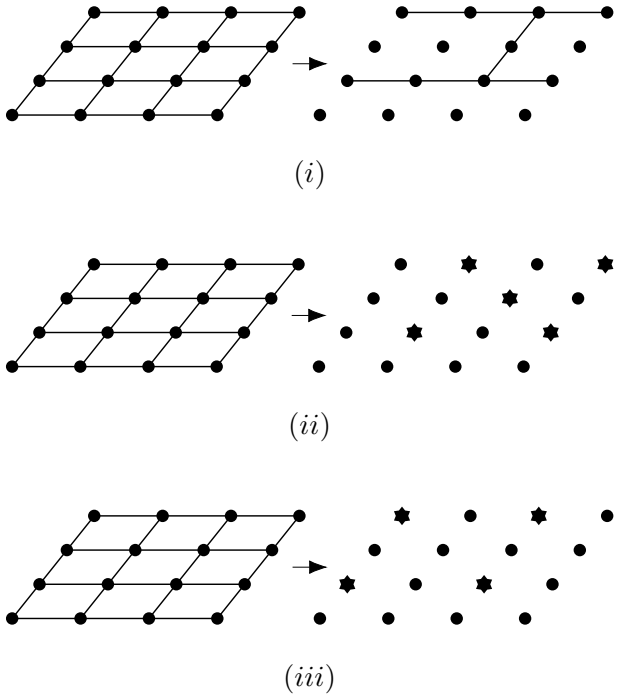

Figure 3: Verifier chooses a random ordering of $2\kappa + 1$ graph states - the computational graph on the right of figure (i); κ identical trap graphs on the right of figure (ii) which have traps (starred nodes) on the even parity positions of the sub-graph that corresponds to the computational graph; and κ identical graphs on the right of figure (iii) which have traps on the odd parity positions of the sub-graph that corresponds to the computational graph. All of these graphs can be generated from a square lattice (on the left) by replacing $|+\rangle$ qubits with $|0\rangle$ at the positions (isolated dot nodes) we do not want entangled with their neighbours when cZ is applied. Further detail on the carving procedure, which can be made blind (Section 3), is provided in Appendix A.

2. Verifier prepares the qubits that compose the cluster state one by one and sends them to the prover.

   (a) The dummy qubits are prepared in $\{|d_i\rangle : d_i \in_R \{0,1\}\}$.

   (b) The rest of the qubits are prepared in $Z^{d_{k\sim j}} |+_{\theta_j}\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i(\theta_j + d_{k\sim j}\pi)} |1\rangle)$, where θ_j is chosen uniformly at random from the set $A = \{0, \frac{\pi}{8}, \frac{2\pi}{8}, \ldots, \frac{15\pi}{8}\}$ and $d_{k\sim j}$ is the parity of the d's of all neighbours of j. Notice that the set A is different from the original trap-based protocol of Ref. [29].

3. Verifier sends the encrypted computational measurement angles to the prover: $\delta_i = \theta_i + (-1)^{r'_i}\phi_i + r_i\pi$ for $r_i, r'_i \in_R \{0,1\}$. Parameters r_i, r'_i create a classical one-time pad on measurement outputs.

4. Prover entangles all received qubits according to the $2\kappa+1$ cluster states, each of dimension $n \times m$, by applying cZ gates for every edge of each cluster.

5. Prover measures all qubits simultaneously in angles δ_i and returns the measurement results to the verifier.

6. Verifier applies a bit flip to the output bit i when $r_i = 1$ and to its (non-dummy) neighbours when $r'_i = 1$ to undo the classical one time pad. The output string $\boldsymbol{x}$ of the measurements of the target computation is the output of the protocol. The verifier sets an extra bit to accept if all the traps give the correct result (decoded measurement result 0).

A variation of the protocol can have the prover to entangle all the graphs directly in the 'extended brickwork state' instead of the square lattice state. This leaks no extra more information to the prover from what is publicly known. However, we seek a more generic prover and keep the protocol as presented.

The resource count of the protocol is as follows. The number of qubits prepared by the verifier and sent to the prover one at a time is $(2\kappa+1)N$ where N is the original size of the computation. The classical information exchanged is linear in N and can be sent in one round. Similarly the classical outcomes of the measurements can be sent in one go. The prover is required to entangle all neighbouring qubits in a square lattice and apply single qubit measurements in the xy-plane.

4.2 Proof of Completeness

To prove completeness we assume that the prover honestly follows the prescribed steps (up to bounded noise). Before considering the noisy case, we show that for the noiseless prover, the fidelity of the target computation and the trap computation to the correct ones are both unity.

We begin with a circuit diagram of the operations on the prover's side in Fig. (4). Any measurement by angle $\{\delta_i\}$ for the prover is mathematically decomposed into a z-rotation (R_z) controlled by δ_i and a Pauli X measurement. Without loss of generality, since everything before the measurements is unitary we can assume that even

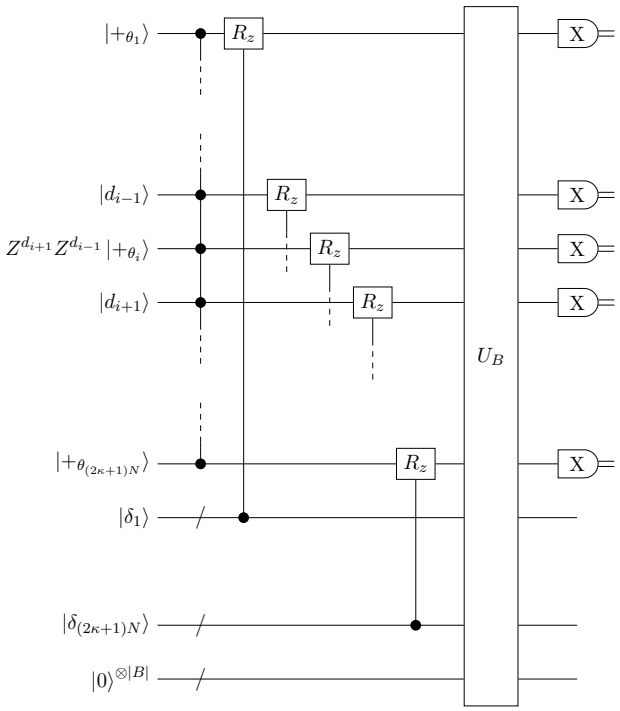

Figure 4: The inputs, other than prover's private system $|0\rangle^{\otimes|B|}$, are the qubits prepared by the verifier in steps 1-3 of Protocol 1, for both target and trap rounds. We represent the prover's operation (steps 4-6 in Protocol 1) upon their receipt. Qubit at position i is a trap qubit surrounded by dummy qubits at positions $i-1$ and $i+1$. U_B is an arbitrary unitary deviation on the prover's system. When prover is honest $U_B = I$.

a dishonest prover will apply the correct unitary operators and then chose his deviation U_B on all systems, including his private qubits $|0\rangle^{\otimes|B|}$. Since we are proving completeness in this section, we assume $U_B = I$. The measurement angles δ_i received by the prover are represented as computational basis multi-qubit states $|\delta_i\rangle$.

The circuit in Fig. (4) can be simplified in a number of ways, resulting in the circuit of Fig. (5). The cZ gates between the dummy qubits and their neighbours cancel the Pauli Z prerotation on the neighbours. Also, we can write explicitly the rotation angles on each of the controlled R_z gates and remove the control lines.

Further simplification follows when the z-rotations by θ_i which are part of the R_z gates and the z-rotations by θ_i applied by the verifier to the qubits before sending them to the prover mutually cancel after commuting with the cZ gates. Notice that the dummy qubits are an exception since θ_i rotations remain but have no effect other than a global phase. Moreover, we can extract the Pauli operators from the R_z by

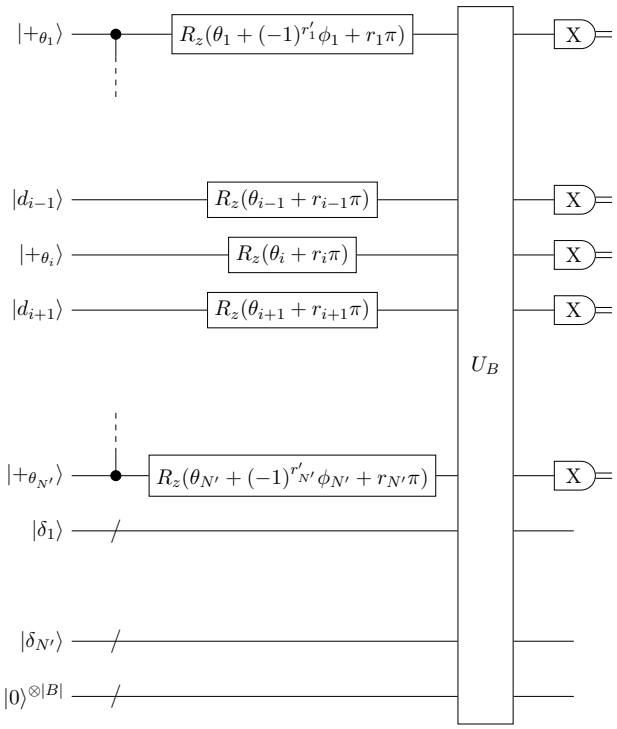

Figure 5: When applying the corresponding entangling operations in Fig. (4), dummy qubits at positions $i-1$ and $i+1$ have the effect of isolating their neighbours and cancelling the neighbours' pre-rotations that depend on parameters d_{i-1}, d_{i+1} (here the only neighbour depicted is the trap qubit at position i). Also, unitary rotations of Fig. (4) are written explicitly. Remember that for dummy and trap qubits angles ϕ take value 0. For clarity of the figure we have used $N' \equiv (2\kappa+1)N$.

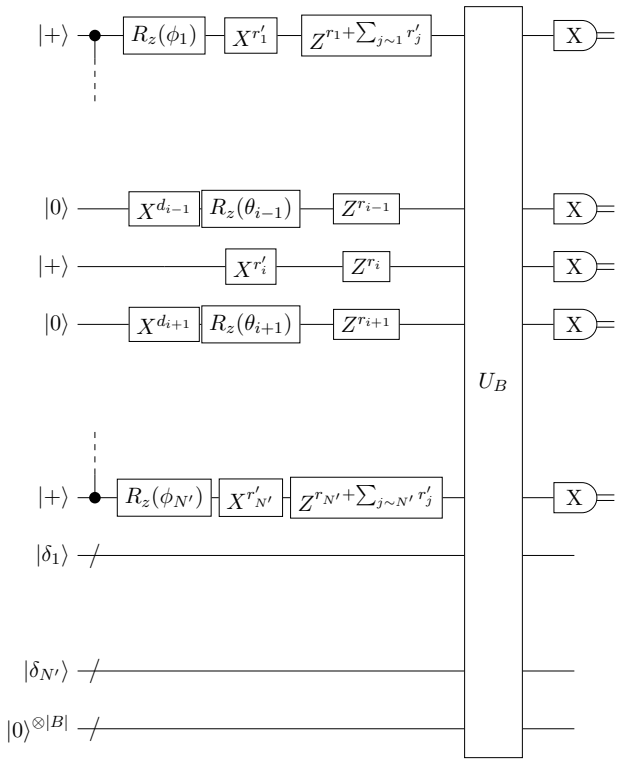

Figure 6: Each z-rotation by θ in Fig. (5) undoes the corresponding pre-rotations of the qubits (except for the dummies that have no pre-rotation by θ). For any qubit k, operations in the form $R_z((-1)^{r'_k}\phi_k)$ in Fig. (5) can be written as $X^{r'_k}R_z(\phi_k)X^{r'_k}$ and the $X^{r'_k}$ before (in temporal order) when commuting with the entanglement operators can be written as $Z^{r'_k}$ on the neighbours (this has an effect on qubits 1 and N' in this figure). All Pauli operators here are written separately from $z-$rotations. Notice that we can write an extra $X^{r'_i}$, with $r'_i \in_R \{0,1\}$, applying on the trap qubit i since $X|+\rangle = |+\rangle$.

applying identities: $R_z(-\chi) = XR_z(\chi)X$ and $R_z(\chi+\pi) = ZR_z(\chi)$. The Pauli X operators from the left hand side of R_z can be rewritten as Z rotations on their entangled neighbours. This results in the circuit diagram depicted in Fig. (6). Notice that the remaining Pauli X operators do not have any effect on the Pauli X measurements (we recall that in this proof $U_B = I$) and the Pauli Z operators flip the measurement results.

Let us denote all the measurement outcomes of the protocol except the dummy qubit measurements by the binary vector $\boldsymbol{x}$ and $p(\boldsymbol{x})$ the probability of obtaining it. Let $q^{\text{exc}}(\boldsymbol{x})$ denote the exact probability of obtaining $\boldsymbol{x}$ in an non-encrypted MBQC implementation using the same measurement pattern $\{\phi\}_i^N$ as input. The only difference between the actual and the non-encrypted case are the Pauli Z operators before the measurements, which flip the outcomes. Therefore, by relabelling the probabilities $p(\boldsymbol{x})$ to $p(\boldsymbol{x'})$, where $x'_i = x_i \oplus r_i \oplus \sum_j r'_{j\sim i}$, we get $q^{\text{exc}}(\boldsymbol{x}) = p(\boldsymbol{x'})$. In other words, in the noiseless case, we can sample from the exact distribution by simply correcting the bit flips caused by the random Pauli Z, which are known to the verifier.

In MBQC, we can also write the distribution in terms of unitaries, labelled by the measurement outcomes (the so-called branches of the MBQC computation) of all the layers except the last. For dimension $m \times n$, we have (up to global phases)

$$q^{\text{exc}}(\boldsymbol{x}) = \frac{|\langle x_{(n-1)m+1},\ldots,x_{nm}| U_{x_1,\ldots,x_{(n-1)m}} |+_1,\ldots,+_m\rangle|^2}{2^{(n-1)m}} \tag{10}$$

since all the computational branches $(x_1,\ldots,x_{(n-1)m})$ are equiprobable and they define a unitary operation on the input [19]. For the trap qubits this distribution is deterministic

since each qubit is prepared in the $|+\rangle$ state, remains isolated throughout the computation and is measured in the $|\pm\rangle$ basis.

Now, consider local bounded noise of the form of Eqn. (9) after every elementary operation j, including preparation, entangling and measurement. The operations that can introduce noise in a single round of the protocol include N preparations at the verifier's end and at most $2N$ entanglements and N measurements at the prover's end. This is an upper bound of $4N$ operations. The fidelity F_c^2 of the noisy output of the target computation to the noiseless one (which is the correct one as we proved above) cannot be smaller than $1-(N\epsilon_V+3N\epsilon_P)$. Since for any two states ρ,σ, $D(\rho,\sigma)\leq\sqrt{1-F^2(\rho,\sigma)}$, this is an upper bound in total variation distance for the target computation $1-\delta=\sqrt{N(\epsilon_V+3\epsilon_P)}$.

Completeness means that our scheme should accept with high probability in the case of bounded noise. The acceptance of the scheme, according to Def. (2), depends on our estimate $\widehat{F}_t^2$ of the acceptance probability of the protocol F_t^2. Given the above bounded noise, F_t^2 cannot be smaller than $1-\kappa N(\epsilon_V+3\epsilon_P)$.

Our estimate for F_t^2 comes from M i.i.d. repetitions of the protocol. By Hoeffding's inequality, repeating $M=\log(1/\beta)/(2\kappa^2N^2(\epsilon_V+\epsilon_P)^2)$ times gets us $\kappa N(\epsilon_V+\epsilon_P)$-close in our estimation with confidence $1-\beta$. In order to have high probability of acceptance we need to set the limit for accepting the estimate to $(1-\kappa(2N\epsilon_V+4N\epsilon_P))$. Then our probability of accepting is as high as our confidence. Setting this limit is necessary to get high completeness but will have an effect in the soundness.

4.3 Proof of Soundness

The proof of soundness of Theorem 1 is based on the fact that the fraction of accepting protocols, $\widehat{F}_t^2$, is a good estimator of a lower bound in the fidelity F_c^2 of the target computation. Thus, looking at $\widehat{F}_t^2$ gives us with high confidence a lower bound on the fidelity, or similarly an upper bound on total variation distance var, as defined in Eq. (1).

We outline the main arguments employed to prove this theorem in stages here, and provide the explicit algebraic derivations in Appendix B.

Firstly, a unitary deviation U_B applied before the measurements, depicted in Fig. (6), captures in all generality the prover's dishonesty. To see this, consider the case when the prover performs measurements different from the honest ones. This corresponds to applying a unitary basis rotation followed by Pauli X measurements. Then, U_B applies also on the prover's private subsystem so he can use this power to replace the qubits he receives with any other qubits he chooses to prepare privately. In any case, he has to report some classical measurement results so we always keep the final Pauli X measurements in the picture. Our proof should therefore apply to any choice of U_B.

Secondly, we bound the total variation distance of the output distribution via the trace distance $D(\rho_c,\rho_c')$, where ρ_c represents the state of the computational system just *after* the Pauli X measurements if the prover is honest and ρ_c' the same state if the prover is dishonest. Thus,

$$\begin{aligned}\text{var}\leq D(\rho_c,\rho_c')&\leq\sqrt{1-F^2(\rho_c,\rho_c')}\\&=\sqrt{1-\text{Tr}^2(\sqrt{\rho_c\rho_c'})}\end{aligned}\quad(11)$$

The main idea leading to the statement of the theorem is that the acceptance probability F_t^2 minus a lower bound on the fidelity of the computational system $F^2(\rho_c,\rho_c')$ is small, when averaged over the random parameters. Therefore, by estimating F_t^2 (by counting the fraction of acceptances over many repetitions of the protocol), we get a good estimate of a lower bound on the fidelity of the computational system and therefore an upper bound on var. We begin our analysis for the case of perfect preparations and subsequently incorporate the effect of noise.

Averaged over the random parameters, the probability of getting all trap outcomes 1, summing over the random variables r_i, r_i', d_i and θ_i, is calculated in Appendix B as

$$F_t^2=\sum_{\boldsymbol{t}}p(\boldsymbol{t})\sum_k|\alpha_k|^2\prod_{i\in\boldsymbol{t}}|\,\langle+|_i\,P_{k|i}\,|+\rangle_i\,|^2,\quad(12)$$

where $\boldsymbol{t}$ is the vector of the indices of the positions of the traps in all 2κ trap systems and $\sum_{\boldsymbol{t}}$ takes all possible values allowed by the construction with equal probability $p(\boldsymbol{t})$. The summation over the random parameters results in the attack on the trap system to be transformed into a convex combination of Pauli operators P_k, each with probability $|\alpha_k|^2$. By $P_{k|i}$ we represent the Pauli operator that applies on qubit i.

The average fidelity F_c of the computational system is

$$F_c \equiv \sum_{\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{t}} p(\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{t}) F(\rho_c, \rho_c') \tag{13}$$

$$= \sum_{\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{t}} p(\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{t}) \mathrm{Tr}(\sqrt{\rho_c \rho_c'}), \tag{14}$$

where ρ_c and ρ_c' represent the honest and dishonest state of the target computation just after the Pauli X measurements. Calculation, presented in detail in Appendix B, leads to

$$F_c^2 \geq \sum_{\boldsymbol{t}} p(\boldsymbol{t}) \sum_k |\alpha_k|^2 \prod_{i \in \boldsymbol{c(t)}} |\langle +|_i P_{k|i} |+\rangle_i|^2 \tag{15}$$

where $\boldsymbol{c(t)}$ denotes the positions of the qubits that participate in the computation and depends on the random ordering of the $2\kappa+1$ rounds and therefore is a function of the position of the traps.

In general we prove that

$$F_t^2 - F_c^2 \leq \frac{\kappa!(\kappa+1)!}{(2\kappa+1)!} \equiv \Delta_\kappa. \tag{16}$$

The verification scheme output bit is set to accept or reject by averaging over M repetitions of the protocol and comparing our estimate of F_t^2 with $(1 - \kappa N(2\epsilon_V + 4\epsilon_P))$ (set by completeness). By Hoeffiding's inequality repeating $M = \log(1/\beta)/(2\kappa^2 N^2(\epsilon_V + \epsilon_P)^2)$ times gets us $\kappa N(\epsilon_V + \epsilon_P)$-close in our estimation with confidence $1-\beta$. Therefore, $F_t^2 - F_c^2 \leq \kappa N(3\epsilon_V + 5\epsilon_P) + \Delta_\kappa$. This means that for the total variation distance we have an upper bound, which gives the soundness parameter ε of Def. (3)

$$\mathrm{var} \leq \varepsilon = \sqrt{\kappa N(3\epsilon_V + 5\epsilon_P) + \Delta_\kappa}.$$

5 Fault-tolerant Verification of Ising Sampler

Ensuring $N\epsilon_V$ and $N\epsilon_P$ in Theorem 1 to be constant will get harder experimentally for increasing N. Therefore, we present two new fault-tolerant verification schemes where the total noise scales linearly with the size, and prove that it provides a distribution that is hard to sample from classically upto constant additive error. We then prove that noise scaling with system size does not prevent us from verifying the prover's distribution with completeness and soundness parameters independent of the problem size.

Quantum fault tolerance strategies such as due to RHG [63] can overcome the challenge of noise scaling with system size. This involves gate distillation requiring adaptive operations which are beyond the Ising sampler. On the target computation, our fault tolerant verification schemes overcome this adaptivity by using arguments for free postselection due to Fujii [32] as applied to the verification of quantum supremacy. On the trap computation, we do not require any adaptivity since we chose it to be Clifford. This keeps our fault tolerant verification schemes within the Ising sampler, allowing verification of quantum supremacy in the presence of total noise scaling linearly with the size. Note that a non-Clifford trap computation would suffer due to nonadaptivity. Time complexity of the quantum operations in the protocol is constant and the number of qubits needed is $O(N\mathrm{PolyLog}(N))$ [63], the polylogarithmic overhead coming from the properties of the topological code and the use of concatenation in the distillation procedure.

The next issue of fault-tolerant thresholds leads to two fault-tolerant versions of the protocol in Fig. (2) and described in detail in Sections 5.1 and 5.2. The first is called Protocol 2a. It employs the full RHG encoding in the traps leading to the threshold of $\epsilon_{\mathrm{thres}} = 0.75\%$ [63], the same threshold as for universal quantum computation. This is worse than the suggested improvements in the noise thresholds for unverified quantum supremacy [32]. However, our next protocol, Protocol 2b, provides $\epsilon_{\mathrm{thres}} = 1.97\%$ for verified quantum supremacy, which is an underestimate because of the analytical treatment and could potentially be improved by numerical simulation as in Ref. [63].

To achieve this threshold, Protocol 2b, replaces error correction with error detection when performing the RHG encoding on the trap qubits. This is possible because the trap qubits are isolated and can be retransmitted individually without affecting the rest of the trap computation. The numerical value is obtained by performing a threshold calculation of applying the RHG encoding in MBQC (Appendix C). A similar procedure was performed in the circuit model by Fujii [32]. The cost of maintaining the same completeness and soundness as in Protocols 1 and 2a is to re-

place κ in Fig. (2) by $M\kappa$, where M is an extra overhead in the number of qubits depending on the code minimal distance d between and around the defects and the noise parameters ϵ_V and ϵ_P. For example, with $d = 2$ and $\epsilon_V = \epsilon_P = \epsilon$ as the following fractions of the noise threshold, we have

ϵ	$\epsilon_{\text{thres}}/20$	$\epsilon_{\text{thres}}/50$	$\epsilon_{\text{thres}}/100$
M	3×10^8	2863	54

Improvement in M may also be possible with judicious braiding or using an alternative topological code.

An additional intricacy needs resolving for both fault-tolerant protocols. Since blindness is an ingredient in our verification scheme, its straightforward application (on the logical level) risks leaking the logical measurement angles in the distillation procedure, where many copies of the same magic state need to be sent. Also, for the distillation procedure to be effective, we need to reveal information about the state distilled. Our stratagem for circumventing this is to apply blindness on the lowest level of MBQC, on which the fault-tolerant construction is based. The traps are applied at the logical MBQC level, since those are the qubits needing protection from noise, as outlined in Fig. (7).

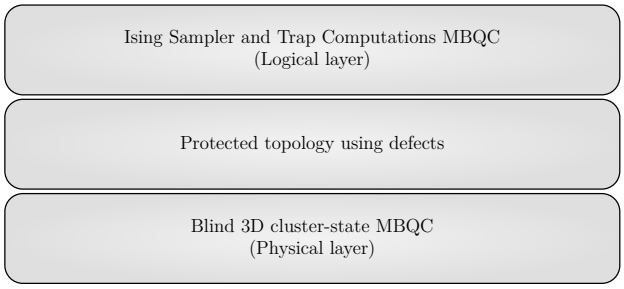

Figure 7: Layered structure of verifiable FT computation.

Our proof of the classical hardness of Ising sampling in this case (Theorem 3) relies on proving the completeness and soundness of verifying conditional probabilities (Theorem 2). They are proved in Section 5.1.

Noise is again of the form of Eqn. (9). Suppose the verifier's noise in each qubit preparation is local, bounded by $\epsilon_V < \epsilon_{\text{thres}}$, the threshold and does not depend on the secret parameters. Assume the honest prover's noise in each elementary operation is bounded by $\epsilon_P < \epsilon_{\text{thres}}$. In order to prove Theorem 2 we make the extra assumption that verifier's noise is independent of the secret parameters.

In the following theorem, ϵ'' is the error rate of the code and scales down exponentially with distance parameter d. Let $q^{\text{nsy}}(\boldsymbol{x}|y=0)$ be the experimental and $q^{\text{exc}}(\boldsymbol{x}|y=0)$ the exact distribution of the Ising sampler, when they are conditioned on the syndrome measurement outcome y giving the null result. The theorem holds for both Protocol 2a and 2b.

Theorem 2 (Fault-tolerant verification scheme)**.** *There exists a verification scheme with Protocol 2a/2b, $M = \log(1/\beta)/(2\epsilon''^2)$ and $l = (1-2\epsilon'')$, that according to Def. (4), which is based on the variation distance between conditional probabilities $q^{nsy}(\boldsymbol{x}|y=0)$ and $q^{exc}(\boldsymbol{x}|y=0)$, is*

$$(1-\beta, 1-\sqrt{\epsilon''}) - \textit{complete}$$

and

$$(1-\beta, \sqrt{3\epsilon'' + \Delta_\kappa}) - \textit{sound}$$

where $\Delta_\kappa = \kappa!(\kappa+1)!/(2\kappa+1)!$.

5.1 Protocol 2a

The fault-tolerant computation scheme used is the one proposed by RHG [63]. Single qubit preparation/distillation, entangling gates (cX) and Pauli X measurements are topologically protected using the three dimensional lattice shown in Fig. (8) and measurement-based implementation of the topological operations (more details on

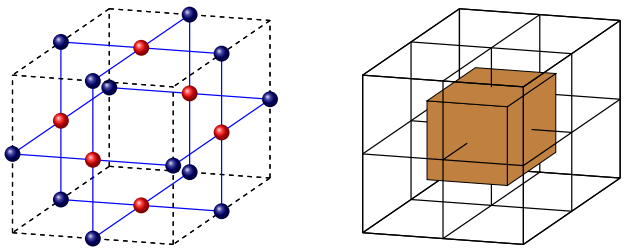

Figure 8: 3D cluster state used in the RHG code using MBQC. Blue dots are the qubits that represent the primal cubic lattice edges (or equivalently the dual cubic lattice faces) and red dots are the qubits that represent the primal cubic lattice faces (or equivalently the dual cubic lattice edges). Entangling operations (cZ) are represented by blue lines. On the right hand side you can see the primal and dual cubes, as are adapted from Refs. [30, 63].

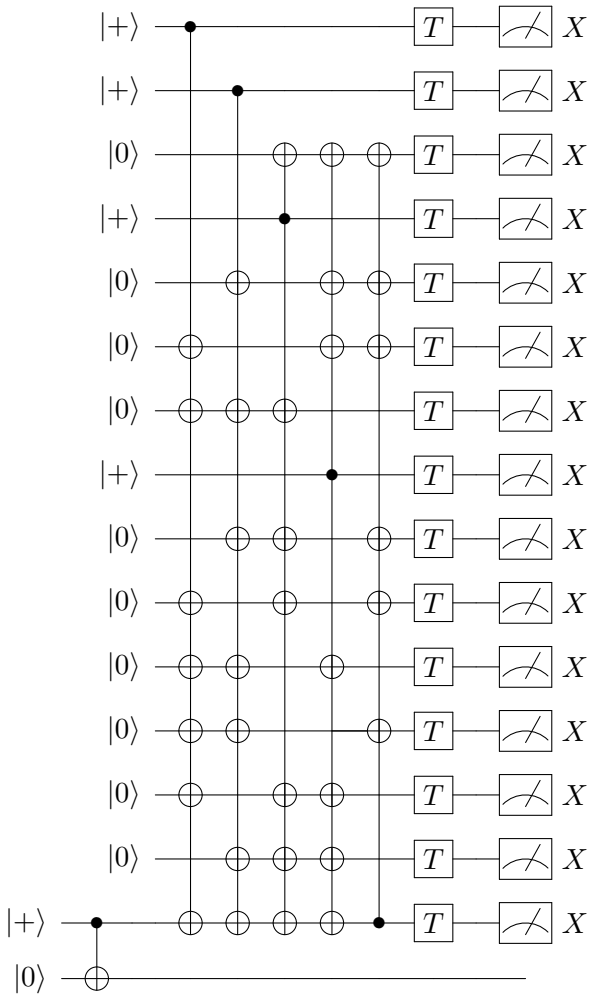

Figure 9: Distillation step [11]. A logical $|+\rangle$ is produced using the $(15, 1, 3)$ quantum Reed-Muller code. Then a transversal T gate applied using the technique of Figure (10), stabilizer measurements and teleportation to an auxiliary qubit gives a 'cleaner' magic state (up to Pauli Z correction on the teleported qubit) when the error syndromes are correct. Everything is topologically protected. Picture adapted from Ref. [34].

the MBQC implementation of the RHG code see [30], [31]). Universality comes from topologically protected concatenated distillation (Fig. (9)) of magic states:

$$|Y\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i\,|1\rangle), |A\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\frac{\pi}{4}}\,|1\rangle)$$

which are generated by single physical qubit measurements. Using the logical distilled magic states and the gate teleportation model one can implement a universal set of gates (Fig. 10). One can simulate an MBQC computation by using these gates and Pauli measurements and consequently add a forth dimension to the system, which comes from the flow of the logical MBQC operations (notice that this layer is distinct from the physical MBQC layer on which the topological code is implemented). The exact usage of RHG encoding in our FT verification scheme depends on whether we use error correction or error detection in our trap computation, giving two separate protocols.

Our first fault-tolerant protocol follows the protocol for the non-fault-tolerant verification of quantum supremacy, introduced in Sec. 4 but using fault tolerance.

Protocol 2a:

1. **Trapification:** Verifier selects a random ordering of $2\kappa + 1$ sufficiently large 3D graphs of Fig. (8), one for the target computation and 2κ for the trap computations. In the target computation round the logical computation is the same as the one in the non-fault-tolerant protocol (see Fig. (1)). In the trap computation rounds, the logical graph contains isolated traps in the same configuration as in the non-fault-tolerant version (see Fig.(3)). We call this the logical layer of our protocol.

2. **Generation of the 'topological code-compatible' circuit:** The above MBQC patterns contain $|\pm_\phi\rangle$ measurements that are not compatible with the topological code. We directly translate the MBQC patterns into a circuit with the same operations, with the difference that the measurements are replaced by teleportation of distilled $|+_\phi\rangle$ states followed by Pauli X measurements. Since our rotations are multiples of $\pi/8$, $|A\rangle$ magic states need be replaced by $\frac{1}{\sqrt{2}}(|0\rangle + e^{i\frac{\pi}{8}}\,|1\rangle)$ states, gate teleportation is as described before (Fig. (10)) with a T gate instead of a S gate correction and distillation based on $(31, 1, 3)$ instead of $(15, 1, 3)$ quantum Reed-Muller code. Adaptive T gates are not applied since we want to keep the model instantaneous - this will be accounted for in our supremacy proof. To avoid adaptivity in distillation we fix which magic states we keep for the next level of distillation independently of the syndrome measurements outcomes - this is not a problem because we have 'free' postselection on the syndrome measurements of the target computation (regis-

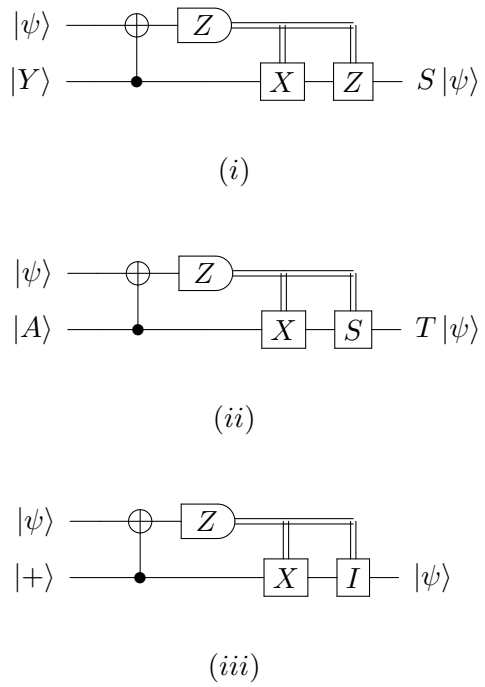

Figure 10: $(i), (ii)$ Gate teleportation (up to global phase) using magic states $|Y\rangle, |A\rangle$, (iii) State teleportation using auxiliary state $|+\rangle$ that mimics gate teleportation (via blindess). These operations are applied in a topologically protected way, both during state distillation using 'impure' states and to implement the corresponding logical operations during computation using distilled states.

ter y).

Even the $|\pm\rangle$ measurements of the traps use inputs that go through the distillation and teleportation procedure (Fig. (10) (iii)). This is in order for the physical attacks to have the same effect on the target and trap computation at the logical level (see proof of verifiability for more).

3. **Topological translation:** The topological translation from the circuit to the topology is straightforward [30].

4. **Blind implementation of topology:** The topological code is implemented at the physical level by MBQC using the 3D-graphs, so that we can implement them blindly using the following encryption.

   (a) Verifier prepares, one by one, the pre-rotated physical qubits $|+_\theta\rangle$, $\theta \in_R \{0, \frac{\pi}{4}, \frac{2\pi}{4}, \ldots, \frac{7\pi}{4}\}$, needed for the blind implementation of the topological protected computation on the three dimensional cluster states and sends them to the prover. Blindness, induced by the random rotations, hides from the prover the physical operations applied and therefore the logical structure of the computation in the topologically protected (vacuum) and isolated qubit region. In particular, the prover is not able to distinguish between implementing distillation and teleportation of a magic state or a $|+\rangle$ state used for computation and testing respectively.

   (b) Verifier sends all the encrypted measurement angles $\delta_i = \theta_i + (-1)^{r'_i}\phi_i + r_i\pi$ for $r_i, r'_i \in_R \{0, 1\}$. Parameters r_i, r'_i are classical one-time pads for the measurement outputs.

   (c) Prover runs the computation, by entangling, measuring all at once and returning the results.

   (d) Verifier classically corrects the returned outcomes using the correction procedure of the quantum error correcting code used in distillation and the topological code and undoes the r, r' pad.

   (e) Verifier accepts the outcome, which is the logical output string $\boldsymbol{x}$ and syndrome measurement bit y ($y = 0$ no error, $y = 1$ error) of the target computation, if all the results of the logical trap computations are correct, otherwise rejects.

Completeness of the protocol follows the same analysis as in the non-fault-tolerant case. We can eliminate the pre-rotations by the θ's of the computation in the lower level MBQC due to θ being in δ and, then, the computation is correct up to Pauli Z corrections before the measurements. Local noise is taken care of by the error correction if it is lower than the threshold of the RHG code. This avoids scaling issues that we had in the non fault-tolerant protocol. In particular because of fault tolerance we get var $\leq \sqrt{\epsilon''}$. Completeness means that our scheme should also accept with high probability and this is achieved by setting the limit to accept the fidelity estimate to $(1 - 2\epsilon'')$. By repeating $N = \log(1/\beta)/(2\epsilon''^2)$ times gets us $\sqrt{\epsilon''}$-close in our estimation with

confidence $1-\beta$. Thus, this is a lower bound on the probability our scheme accepts in the case of completeness.

The proof of soundness is similar to the non fault-tolerant case since the noise can be considered part of the prover's attack that has the same effect on the target computation and the trap at a logical level. We show this in Appendix D.

The threshold of this protocol is the same as the threshold of the RHG code since error correction is used in the trap rounds.

5.2 Protocol 2b

We now adapt Protocol 2a to work with error detection and attain a better threshold. The main idea is that because the traps are isolated qubits one can look at the syndrome measurements of all trap computations and pick from each computation only the logical trap qubit measurements that are correct individually.

The traps in this case test topologically protected qubits of the graph state that implements the target computation *together with the distillation*. This is because we want to have smaller traps, in terms of number of physical qubits, compared to Protocol 2a where a trap can be as large as a magic state distillation circuit. This is crucial because by employing error detection and retransmission, one needs to resend one logical trap every time at least one syndrome measurement in the topologically protected region around the trap fails (we limit this overhead to M times).

To avoid the trap computation being distinguishable from the target, we implement the traps as if all qubits have an injected singular qubit, but the injected singular qubit is prepared in the $|+\rangle$ state and therefore the logical input remains the logical $|+\rangle$. The underlying MBQC blindness hides the qubit that is injected. Each trap computation is now performed M times.

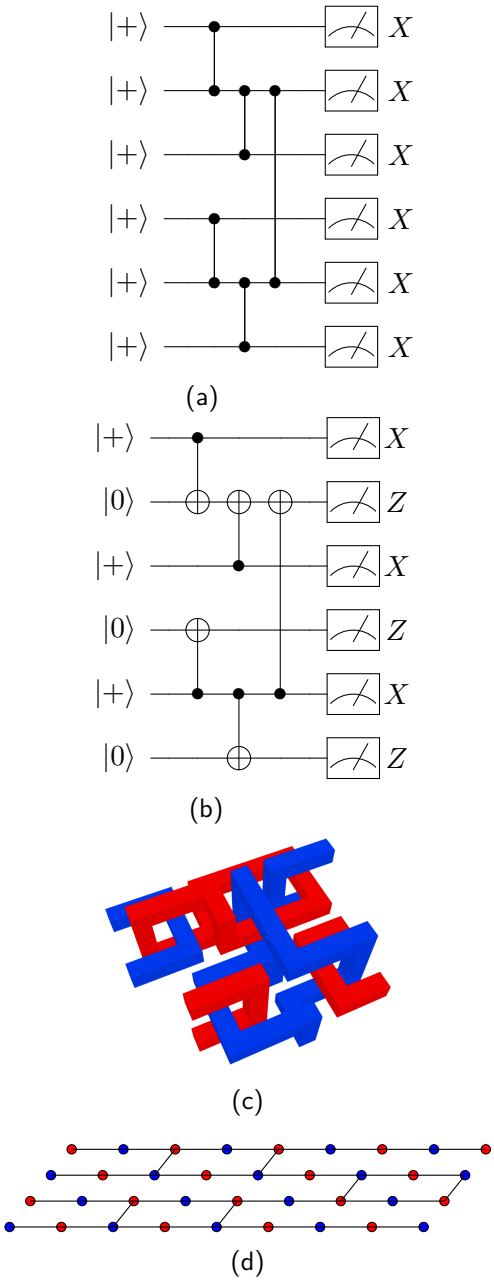

Figure 11: (a) 'H'-shaped building component of brickwork state. (b) Same with cNOT gates where the control is always $|+\rangle$ and the target always $|0\rangle$. (c) Translation into prime (blue)/dual (red) topologically protected qubits. (d) Prime/dual colouring of the topologically protected brickwork state.

Protocol 2b:

1. **Generation of the 'topological code-compatible' MBQC pattern:** Verifier selects a random ordering of $2\kappa M+1$ sufficiently large 3D graphs of Fig. (8), one for the target computation and $2\kappa M$ for the trap computations. For the target computation round: The Ising sampler MBQC pattern of Fig. (1) is translated into a 'topological code-compatible' one, i.e. an MBQC pattern where qubits are prepared as $|+_{k\pi/8}\rangle$ states and always measured in the $\{|\pm\rangle\}$ basis. This translation is possible using again circuits similar to Fig. (10). Notice that this introduces some adaptive T gates that we cannot perform if we want to keep the model instantaneous - this will be accounted for in our supremacy proof. Moreover, topologi-

cal protection requires the distillation of the magic states and this can also be translated into an MBQC pattern. To avoid adaptivity we fix which magic states we keep for the next level of distillation independently of the syndrome measurements outcomes - this is not a problem because we have 'free' post-selection on the target computation (register y). The final MBQC pattern can be also standardised to the form of a brickwork state so that it can be trapified shown as in Fig. (3).

2. **Trapification:** The target computation is the MBQC pattern generated in the previous step. For the trap round, as shown in Fig. (3), we have two types of trap computations by isolating qubits of the brickwork state. This is also 'topological code-compatible'. Qubits are prepared in the $|+\rangle$ or $|0\rangle$ state and are measured in the $\{|\pm\rangle\}$ basis. Notice that in the trap rounds there is no adaptivity. We call this the logical layer of our protocol.

3. **Topological translation:** As shown in Fig. (11) one can translate the 'topological code-compatible' MBQC pattern into a topology that conforms with the topological code. To avoid leaking any information concerning when magic states or dummy qubits are injected, we inject a physical qubit at every logical qubit. Thus, we use the same topology to inject $|+\rangle$ (which is equivalent to not injecting anything in the topology of Fig. (11)) or $|0\rangle$ or a magic state when needed.

4. **Blind implementation of topology:** In order to implement the above topology blindly, so that the prover does not know which physical states we inject, we chose to implement it on MBQC and use the following encryption.

   (a) Verifier prepares, one by one, the pre-rotated physical qubits $|+_\theta\rangle$, $\theta \in_R \{0, \frac{\pi}{4}, \frac{2\pi}{4}, \ldots, \frac{7\pi}{4}\}$

   (b) Verifier sends all the encrypted measurement angles $\delta_i = \theta_i + (-1)^{r'_i}\phi_i + r_i\pi$ for $r_i, r'_i \in_R \{0,1\}$. Parameters r_i, r'_i are classical one-time pads for the measurement outputs.

   (c) Prover runs the computation, by entangling, measuring all at once and returning the results.

   (d) Verifier classically detects the errors in the returned syndrome measurements of the trap computations after undoing the r, r' pad. From the set of the κM logical trap qubits corresponding to each position in the trap graph it selects κ correct ones. This is possible, on average, if M is large enough as described at the end of Appendix C. This results in the quantity Δ_κ in Theorem 2 being averaged over the noise distribution.

   (e) Verifier accepts the outcome, which is the logical output string $\boldsymbol{x}$ and syndrome measurement bit y ($y = 0$ no error, $y = 1$ error) of the target computation, if all the results of the logical trap computations are correct, otherwise rejects.

The proof of correctness and soundness are identical to Protocol 2a.

6 Noisy Computational Supremacy

Assuming the following conjectures, the quantum computational supremacy theorem (Theorem 3) for the noisy case holds.

Conjecture 1 (Average-case hardness)**.** *For* $0 \leq \alpha_1, \beta_1 \leq 1$, *approximating the probability distribution of the Ising sampler by* $p^{\text{apx}}(\boldsymbol{x}|y=0)$ *up to multiplicative error*

$$|p^{\text{apx}}(\boldsymbol{x}|y=0) - q^{\text{exc}}(\boldsymbol{x}|y=0)| \leq \alpha_1 q^{\text{exc}}(\boldsymbol{x}|y=0)$$

in time poly$(|\boldsymbol{x}|, 1/\alpha_1, 1/\beta_1)$ *is #P-hard for at least a fraction* β_1 *of* $\boldsymbol{x}$ *instances.*

Conjecture 2 (Anti-concentration)**.** *There exist some* $0 \leq \alpha_2, \beta_2 \leq 1$, $1/\alpha_2 \in \text{poly}(1/\beta_2)$ *such that for all* x

$$\text{prob}\left(q^{\text{exc}}(\boldsymbol{x}|y=0) \geq \frac{\alpha_2}{2^N}\right) \geq \beta_2 \qquad (17)$$

The above encapsulate two properties for the Ising sampler: the worst to average case hardness equivalence for multiplicative approximations and the probability anti-concentration conjecture.

Theorem 3 (Fault-tolerant hardness)**.** *Assume that Conjectures 1 and 2 hold. Then sampling from the output distribution of the experimental Ising sampler $q^{\text{nsy}}(\boldsymbol{x}, y)$ with a classical machine, assuming a $(\varepsilon', \varepsilon)$-sound verification scheme (Def. 3/Def. 4) accepts with*

$$\varepsilon \leq \frac{(\beta_1 + \beta_2 - 1 - 2^{-N})\alpha_1\alpha_2}{2}, \qquad (18)$$

implies, with confidence ε', a collapse in the polynomial hierarchy to the third level.

In order to have a scheme with positive soundness parameter ε, we need our conjectures to satisfy $\beta_1 + \beta_2 - 2^{-N} \geq 1$.

The proof uses Stockmeyer's theorem [69], which is based on a hypothetical machine and predicts, if classical sampling is possible, the collapse of the polynomial hierarchy to the third level. The collapse of the infinite polynomial hierarchy at any level is considered a highly unlikely event in computational complexity theory.

6.1 Proof

Compared the the FT hardness proof of [32], our proof is for the more general case of additive as opposed to multiplicative approximation, thus answering an open question of that paper.

We follow a similar line of reasoning as the original translationally invariant Ising sampler [36] - proof by contradiction. The main difference is that we use probabilities conditioned on null syndromes. Other differences include adding explanation of intermediate steps, a discussion about obfuscation and breaking the original single Conjecture into two separate ones: one for anti-concentration and one for average case hardness.

We also follow a line similar to an earlier result [12]. Compared to that result our proof does not assign specific numbers to the parameters of the conjectures, but states them in a parametrised fashion.

The following proof holds for a verification scheme according to Def. 4. In the case of a verification scheme of Def. 3 the same proof holds, replacing var^{Post} with var and setting $q^{\text{nsy}}(y = 0) = 1$.

Proof of Theorem 3. If we can classically sample from $q^{\text{nsy}}(\boldsymbol{x}, y)$ (which means that our quantum computer could be a classical impostor), then estimating the probabilities of the distribution with exponential accuracy is in $\#P$: We can construct a polynomial time non-deterministic Turing machine that uses the sampler as an oracle that accepts when a specific string is sampled, so that the probability of that event could be estimated, if we could count the accepting branches. We could also estimate the marginal probabilities $q^{\text{nsy}}(y)$ in such a manner. Notice that we could not apply the same argument for the quantum sampler since we cannot extract its randomness as input to build the oracle. From Stockmeyer's theorem [69], there exists an $FBPP^{NP}$ machine that can compute explicitly the values $p^{\text{apx}}(\boldsymbol{x}, y)$, such that for every $\boldsymbol{x}, y$

$$|p^{\text{apx}}(\boldsymbol{x}, y) - q^{\text{nsy}}(\boldsymbol{x}, y)| \leq \frac{q^{\text{nsy}}(\boldsymbol{x}, y)}{\text{poly}(N)}. \qquad (19)$$

The same can be applied in calculating the marginals. Thus a $FBPP^{NP}$ machine can calculate $q^{\text{nsy}}(y = 0)$, the probability of accepting the syndrome measurements, with accuracy of the same scaling as the joint probability.

Using the fact that $q^{\text{nsy}}(y = 0)$ is non-zero (it is lower bounded by $(1 - \epsilon)^N$, so one can get a non-zero estimate in $\#P$ and approximate it using Stockmeyer) it is easy to prove that for conditional probabilities,

$$|p^{\text{apx}}(\boldsymbol{x}|y = 0) - q^{\text{nsy}}(\boldsymbol{x}|y = 0)| \leq \frac{q^{\text{nsy}}(\boldsymbol{x}|y = 0)}{\text{poly}(N)}. \qquad (20)$$

Applying the triangle inequality, for every $\boldsymbol{x}$ the distance between the values $p^{\text{apx}}(\boldsymbol{x}|y = 0)$ and the exact conditional probability $q^{\text{exc}}(\boldsymbol{x}|y = 0)$ of the Ising sampler is

$$\begin{aligned} &|p^{\text{apx}}(\boldsymbol{x}|y = 0) - q^{\text{exc}}(\boldsymbol{x}|y = 0)| \\ &\leq |p^{\text{apx}}(\boldsymbol{x}|y = 0) - q^{\text{nsy}}(\boldsymbol{x}|y = 0)| \\ &\quad + |q^{\text{nsy}}(\boldsymbol{x}|y = 0) - q^{\text{exc}}(\boldsymbol{x}|y = 0)| \qquad (21) \\ &\leq \frac{q^{\text{nsy}}(\boldsymbol{x}|y = 0)}{\text{poly}(N)} + |q^{\text{nsy}}(\boldsymbol{x}|y = 0) \\ &\quad - q^{\text{exc}}(\boldsymbol{x}|y = 0)|. \qquad (22) \end{aligned}$$

Assuming an $(\varepsilon', \varepsilon)$-sound verification scheme has accepted, it follows that $|q^{\text{nsy}}(\boldsymbol{x}|y = 0) - q^{\text{exc}}(\boldsymbol{x}|y = 0)| \leq 2\varepsilon$, with confidence ε'.

Obfuscation of the probability estimated in this model comes by construction. We can pick a computational branch $(x_1, \ldots, x_{m(n-1)})$ and final layer output string $(x_n, \ldots, x_{mn})$ to estimate at random, without revealing any information to

the sampler. This is possible because the uniform distribution over branches (see Eq. (10) in Section 4.2) is created within a fixed instance of the Ising sampler, with no extra input provided to the sampler. The expectation of $\mathrm{var}^{\mathrm{Post}}$ over the uniform distribution on $\boldsymbol{x}$ is $\leq \frac{2\varepsilon}{2^{mn}}$, where m, n are the dimensions of the logical 'extended' brickwork state and $N = mn$.

Markov inequality relates the probability of a random variable exceeding a certain value with its expectation. For a random variable X and $\gamma > 0$,

$$\mathrm{prob}(X \geq \gamma) \leq \frac{E(X)}{\gamma}. \tag{23}$$

Applying the Markov inequality to the second term in Eqn. (22)

$$\mathrm{prob}(|q^{\mathrm{nsy}}(\boldsymbol{x}|y=0) - q^{\mathrm{exc}}(\boldsymbol{x}|y=0)| \geq \gamma) \leq \frac{2\varepsilon}{2^N\gamma}. \tag{24}$$

By changing variables

$$\mathrm{prob}\left(|q^{\mathrm{nsy}}(\boldsymbol{x}|y=0) - q^{\mathrm{exc}}(\boldsymbol{x}|y=0)| \geq \frac{2\varepsilon}{2^N\gamma}\right) \leq \gamma \tag{25}$$

or

$$\mathrm{prob}\left(|q^{\mathrm{nsy}}(\boldsymbol{x}|y=0) - q^{\mathrm{exc}}(\boldsymbol{x}|y=0)| \leq \frac{2\varepsilon}{2^N\gamma}\right) \geq 1-\gamma. \tag{26}$$

Condensing this with Eqn. (22),

$$\mathrm{prob}\left(|p^{\mathrm{apx}}(\boldsymbol{x}|y=0) - q^{\mathrm{exc}}(\boldsymbol{x}|y=0)| \leq \frac{q^{\mathrm{exc}}(\boldsymbol{x}|y=0)}{\mathrm{poly}(N)} + \frac{2\varepsilon(1+o(1))}{2^N\gamma}\right) \geq 1-\gamma, \tag{27}$$

Thus, for more than $1-\gamma$ fraction of instances of $\boldsymbol{x}$

$$|p^{\mathrm{apx}}(\boldsymbol{x}|y=0) - q^{\mathrm{exc}}(\boldsymbol{x}|y=0)| \leq \frac{q^{\mathrm{exc}}(\boldsymbol{x}|y=0)}{\mathrm{poly}(N)} + \frac{2\varepsilon(1+o(1))}{2^N\gamma}, \tag{28}$$

which means that strongly simulating, i.e. calculating the probabilities, of the Ising distribution with the above mixture of additive and multiplicative accuracy for more than $1-\gamma$ fraction of instances of $\boldsymbol{x}$ is in the third level of the polynomial hierarchy.

We use the two conjectures to continue our proof. From Eqn. (28) and Conjecture 2, setting $\varepsilon_1 \in \frac{2\varepsilon(1+o(1))}{\gamma\alpha_2}$, there must be at least $\beta_2 - \gamma$ fraction of instances of $\boldsymbol{x}$ (we assume $\gamma < \beta_2$) such that

$$|p^{\mathrm{apx}}(\boldsymbol{x}|y=0) - q^{\mathrm{exc}}(\boldsymbol{x}|y=0)| \leq \frac{q^{\mathrm{exc}}(\boldsymbol{x}|y=0)}{\mathrm{poly}(N)} + \varepsilon_1 q^{\mathrm{exc}}(\boldsymbol{x}|y=0) \tag{29}$$

$$\leq (o(1) + \varepsilon_1) q^{\mathrm{exc}}(\boldsymbol{x}|y=0). \tag{30}$$

Let Conjecture 1 hold with $\beta_1 \geq 1 - (\beta_2 - \gamma) + 2^{-N}$ and $\alpha_1 \in o(1) + \varepsilon_1$. These imply for the soundness parameter

$$\varepsilon \leq \frac{(\beta_1 + \beta_2 - 1 - 2^{-N})\alpha_2\alpha_1}{2}, \tag{31}$$

which is positive for $\beta_1 + \beta_2 - 2^{-N} \geq 1$.

Then, there exists at least one instance for which the multiplicative approximation the $FBPP^{NP}$ Stockmeyer machine have calculated is $\#P$-hard.

Then, $PH \subseteq P^{\#P} \subseteq P^{FBPP^{NP}} \subseteq \Sigma_3^P$, where the first inclusion is given by Toda's theorem, and the polynomial hierarchy collapses to the third level, an event expected to be highly unlikely.

□

Our average case conjecture is implicitly contained in a stronger conjecture that includes anticoncetration in [36]. Notice that our average case conjecture, however, applies to a slightly different distribution. The difference is that, in our case, we can have extra rotations on some of the measurement angles, as a consequence of not applying the correction gates conditioned on the measurement outcome of the teleportation step of the FT gates (Fig. (10) (ii)). This issue will affect the implementation of the measurements with nonzero angles in the extended brickwork state (Figure 1 (ii)). Assuming magic $|+_{\pi/8}\rangle$ states are used for the implementation of the $\pi/8$ rotations ($\sqrt{T}$ gates), the byproduct is a $k_0\pi/4$ rotation on the measurements of the brickwork state vertices (Fig. (1) (i)), for some k_0 which depends on the measurement outcomes of the magic state teleportation steps. The original argument made in [36] to support their average case conjecture relies on the fact that from random measurement outcomes of the vertices of the extended brickwork state, a uniform rotation over the 8 different $\{k\pi/4\}$ angles is produced on the brickwork state.

The argument is that this corresponds to random circuits which will likely produce highly entangled states (see also p. 7 in Supplemental Material of Ref. [36]). In our case it is the same, because the extra rotations cannot change the uniformity of these angles. Thus, the computation applied is based on a random brickwork MBQC pattern and connections to the random circuit model, such as in [9], can be made. In the latter paper, the average hardness of sampling from a random circuit is supported by drawing connections to quantum chaos and some numerical evidence. In another recent result ([10]), average-case hardness has been proven for random circuits for the exact case.

7 Discussions

Quantum computational supremacy demonstration is believed to be easier than universal quantum computing since it may not have to fulfil atleast one of DiVincenzo's criteria. Our work shows that fault-tolerant verifiable quantum supremacy is quantitatively easier than fault-tolerant universal quantum computation in terms of thresholds. This relies on combining the notion of post-selected thresholds with trap-based verification which allows error-detection-based fault tolerance to combat noise. Such a combination is not known to exist for other quantum verification methods. In the trap-based verification schemes we use, it is the isolatable nature of the traps that enables error-detection-based fault tolerance.

The techniques developed here have a wide range of applicability. We apply it to the Ising sampler as a specific example of a model for quantum supremacy. For example, they could be applied in implementations of the Boson Sampling model [1] in a fault-tolerant quantum computer based on qubits [61]. Our methods should also apply to the random circuit IQP model [14] which, in the 'graph program' implementation [67] requires a smaller than the Ising sampler, but non-planar, resource state. Finally, it can be applied to recently studied quantum supremacy architectures on low-periodicity planar lattices [7]. The only requirement for our isolated trap computation technique is that the underlying graph state is bipartite. Thus, it can even be used to simplify the original verification protocol [29] for a universal quantum computer, in the case it runs a classical output problem, and use our technique to implement it in a fault-tolerant way.

Trap-based techniques require blindness, which is not believed possible with a classical verifier [3, 59]. Even verification protocols that do not require blindness, such as [4], still need some level of quantum encryption. This is true for any protocol based on quantum authentication schemes [6], made possible by a quantum verifier. Verification protocols with classical verifiers exist [52, 65, 67], but require extra assumptions such as additional computational hardness conjectures or non-communicating provers respectively. For general reviews of blind and verifiable protocols see Refs. [28, 38].

Our work is one of the first on fault-tolerant verification, which was known to be a challenging open question. Another recent progress [33] presents a fault-tolerant verification technique for universal MBQC that requires the verifier to perform measurements, as opposed to preparations as in our scheme. Our scheme is complementary to contemperaneous work on composable verification of IQP, which is a classical hypothesis test with the verifier preparing perfect stabilizer states and the prover using a non-planar graph [57]. A formal proof of composability of our protocol is a desirable next step and may be developed using the methods given in [24].

A direction for future investigation should be the potential of other known fault-tolerant quantum codes in providing improved post-selected thresholds, as well as the search for quantum codes for non-universal models. Another direction should be the study of known quantum supremacy models for which a verifiable version using the same restricted physical assumptions as the original exists [45], as well as the development of such new non-universal models. More technically, an open problem for our verification scheme is to find a graph state with local rotations being only multiples of $\pi/4$ and still generate random universal logical measurement angles as in the existing scheme. This will make the fault-tolerant version easier because it will be based on more standard magic states. Also, other universal constructions with xy-plane measurements can also be considered [53, 55].

Acknowledgements

We thank Zhang Jiang, Ashley Montanaro, Michael Bremner for early discussions and Elham Kashefi, Petros Wallden and Andru Gheorghiu for discussions on the trap-based verification technique. We thank the anonymous referees for valuable feedback that helped improve the manuscript. We thank Dominic Branford and Samuele Ferracin for discussions and the former for also helping with the figures. This work was supported, in part, by the UK EPSRC (EP/K04057X/2), and the UK National Quantum Technologies Programme (EP/M013243/1).

References

[1] Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. *Theory of Computing*, 9(1):143–252, 2013. ISSN 1557-2862. DOI: 10.4086/toc.2013.v009a004.

[2] Scott Aaronson and Lijie Chen. Complexity-theoretic foundations of quantum supremacy experiments. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 79. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017. DOI: 10.4230/LIPIcs.CCC.2017.22.

[3] Scott Aaronson, Alexandru Cojocaru, Alexandru Gheorghiu, and Elham Kashefi. On the implausibility of classical client blind quantum computing. *arXiv preprint arXiv:1704.08482*, 2017. URL https://arxiv.org/abs/1704.08482.

[4] Dorit Aharonov, Michael Ben-Or, and Elad Eban. Interactive proofs for quantum computations. In *Proceedings of Innovations in Computer Science 2010*, ICS2010, pages 453–, 2010. URL http://conference.iiis.tsinghua.edu.cn/ICS2010/content/paper/Paper_35.pdf.

[5] Leandro Aolita, Christian Gogolin, Martin Kliesch, and Jens Eisert. Reliable quantum certification of photonic state preparations. *Nature communications*, 6, 2015. DOI: 10.1038/ncomms9498.

[6] Howard Barnum, Claude Crepeau, Daniel Gottesman, Adam Smith, and Alain Tapp. Authentication of quantum messages. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 449–458. IEEE, 2002. DOI: 10.1109/SFCS.2002.1181969.

[7] Juan Bermejo-Vega, Dominik Hangleiter, Martin Schwarz, Robert Raussendorf, and Jens Eisert. Architectures for quantum simulation showing a quantum speedup. *Phys. Rev. X*, 8:021010, Apr 2018. DOI: 10.1103/PhysRevX.8.021010.

[8] K. Binder and A. P. Young. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Rev. Mod. Phys.*, 58:801–976, Oct 1986. DOI: 10.1103/RevModPhys.58.801.

[9] Sergio Boixo, Sergei V Isakov, Vadim N Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J Bremner, John M Martinis, and Hartmut Neven. Characterizing quantum supremacy in near-term devices. *Nature Physics*, 14(6):595, 2018. DOI: 10.1038/s41567-018-0124-x.

[10] Adam Bouland, Bill Fefferman, Chinmay Nirkhe, and Umesh Vazirani. " quantum supremacy" and the complexity of random circuit sampling. In *10th Innovations in Theoretical Computer Science Conference (ITCS 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018. DOI: 10.4230/LIPIcs.ITCS.2019.15.

[11] Sergey Bravyi and Alexei Kitaev. Universal quantum computation with ideal clifford gates and noisy ancillas. *Phys. Rev. A*, 71:022316, Feb 2005. DOI: 10.1103/PhysRevA.71.022316.

[12] Michael Bremner, Ashley Montanaro, and Daniel Shepherd. Average-case complexity versus approximate simulation of commuting quantum computations. *Physical Review Letters*, 117, 8 2016. ISSN 0031-9007. DOI: 10.1103/PhysRevLett.117.080501.

[13] Michael J. Bremner, Richard Jozsa, and Dan J. Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467 (2126):459–472, feb 2011. ISSN 1364-5021. DOI: 10.1098/rspa.2010.0301.

[14] Michael J. Bremner, Ashley Montanaro, and Dan J. Shepherd. Average-case complexity versus approximate simulation of commut-

ing quantum computations. *Phys. Rev. Lett.*, 117:080501, Aug 2016. DOI: 10.1103/PhysRevLett.117.080501.

[15] Michael J. Bremner, Ashley Montanaro, and Dan J. Shepherd. Achieving quantum supremacy with sparse and noisy commuting quantum computations. *Quantum*, 1:8, April 2017. ISSN 2521-327X. DOI: 10.22331/q-2017-04-25-8.

[16] Anne Broadbent. How to verify a quantum computation. *Theory of Computing*, 14(11):1–37, 2018. DOI: 10.4086/toc.2018.v014a011.

[17] Anne Broadbent, Joseph Fitzsimons, and Elham Kashefi. Universal blind quantum computation. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 517–526. IEEE, 2009. DOI: 10.1109/FOCS.2009.36.

[18] Iulia Buluta and Franco Nori. Quantum simulators. *Science*, 326(5949):108–111, 2009. ISSN 0036-8075. DOI: 10.1126/science.1177838.

[19] Andrew M Childs, Debbie W Leung, and Michael A Nielsen. Unified derivations of measurement-based schemes for quantum computation. *Physical Review A*, 71(3):032318, 2005. DOI: 10.1103/PhysRevA.71.032318.

[20] Marcus Cramer, Martin B Plenio, Steven T Flammia, Rolando Somma, David Gross, Stephen D Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. Efficient quantum state tomography. *Nature communications*, 1:149, 2010. DOI: 10.1038/ncomms1147.

[21] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A*, 80:012304, Jul 2009. DOI: 10.1103/PhysRevA.80.012304.

[22] Eric Dennis, Alexei Kitaev, Andrew Landahl, and John Preskill. Topological quantum memory. *Journal of Mathematical Physics*, 43(9):4452–4505, 2002. DOI: 10.1063/1.1499754.

[23] David P. DiVincenzo. The physical implementation of quantum computation. *Fortschritte der Physik*, 48(9-11):771–783, 2000. ISSN 1521-3978. DOI: 10.1002/1521-3978(200009)48:9/11<771::AID-PROP771>3.0.CO;2-E.

[24] Vedran Dunjko, Joseph F Fitzsimons, Christopher Portmann, and Renato Renner. Composable security of delegated quantum computation. In *Advances in Cryptology–ASIACRYPT 2014*, pages 406–425. Springer, 2014. DOI: 10.1007/978-3-662-45608-8_22.

[25] Eddie Farhi. Quantum supremacy through the quantum approximate optimization algorithm. In *APS Meeting Abstracts*, 2017. URL `http://adsabs.harvard.edu/abs/2017APS..MARA19004F`.

[26] Samuele Ferracin, Theodoros Kapourniotis, and Animesh Datta. Reducing resources for verification of quantum computations. *Phys. Rev. A*, 98:022323, Aug 2018. DOI: 10.1103/PhysRevA.98.022323.

[27] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, 1982. ISSN 1572-9575. DOI: 10.1007/BF02650179.

[28] Joseph F Fitzsimons. Private quantum computation: an introduction to blind quantum computing and related protocols. *npj Quantum Information*, 3(1):23, 2017. DOI: 10.1038/s41534-017-0025-3.

[29] Joseph F Fitzsimons and Elham Kashefi. Unconditionally verifiable blind quantum computation. *Physical Review A*, 96 (1):012303, 2017. DOI: 10.1103/PhysRevA.96.012303.

[30] Austin G Fowler and Kovid Goyal. Topological cluster state quantum computing. *Quantum Information and Computation*, 9(9-10): 0721–0738, 2009. DOI: 10.26421/QIC9.9-10.

[31] Keisuke Fujii. *Quantum Computation with Topological Codes: from qubit to topological fault-tolerance*, volume 8. Springer, 2015. DOI: 10.1007/978-981-287-996-7.

[32] Keisuke Fujii. Noise threshold of quantum supremacy. 2016. URL `https://arxiv.org/abs/1610.03632`.

[33] Keisuke Fujii and Masahito Hayashi. Verifiable fault tolerance in measurement-based quantum computation. *Phys. Rev. A*, 96: 030301, Sep 2017. DOI: 10.1103/PhysRevA.96.030301.

[34] Keisuke Fujii and Tomoyuki Morimae. Commuting quantum circuits and complexity

of ising partition functions. *New Journal of Physics*, 19(3):033003, 2017. DOI: 10.1088/1367-2630/aa5fdb.

[35] Takeshi Fukuhara, Adrian Kantian, Manuel Endres, Marc Cheneau, Peter Schausz, Sebastian Hild, David Bellem, Ulrich Schollwock, Thierry Giamarchi, Christian Gross, Immanuel Bloch, and Stefan Kuhr. Quantum dynamics of a mobile spin impurity. *Nat Phys*, 9(4):235–241, apr 2013. ISSN 1745-2473. DOI: 10.1038/nphys2561.

[36] Xun Gao, Sheng-Tao Wang, and L.-M. Duan. Quantum Supremacy for Simulating a Translation-Invariant Ising Spin Model. *Physical Review Letters*, 118(4):040502, jan 2017. ISSN 0031-9007. DOI: 10.1103/PhysRevLett.118.040502.

[37] I. M. Georgescu, S. Ashhab, and Franco Nori. Quantum simulation. *Rev. Mod. Phys.*, 86:153–185, Mar 2014. DOI: 10.1103/RevModPhys.86.153.

[38] Alexandru Gheorghiu, Theodoros Kapourniotis, and Elham Kashefi. Verification of quantum computation: An overview of existing approaches. *Theory of computing systems*, pages 1–94, 2018. DOI: 10.1007/s00224-018-9872-3.

[39] Leslie Ann Goldberg and Heng Guo. The complexity of approximating complex-valued ising and tutte partition functions. *Computational Complexity*, 26(4):765–833, 2017. DOI: 10.1007/s00037-017-0162-2.

[40] D Hangleiter, M Kliesch, M Schwarz, and J Eisert. Direct certification of a class of quantum simulations. *Quantum Science and Technology*, 2(1):015004, mar 2017. ISSN 2058-9565. DOI: 10.1088/2058-9565/2/1/015004.

[41] Aram W. Harrow and Ashley Montanaro. Quantum computational supremacy. *Nature*, 549(7671):203–209, sep 2017. ISSN 0028-0836. DOI: 10.1038/nature23458.

[42] Masahito Hayashi and Tomoyuki Morimae. Verifiable measurement-only blind quantum computing with stabilizer testing. *Phys. Rev. Lett.*, 115:220502, Nov 2015. DOI: 10.1103/PhysRevLett.115.220502.

[43] M. Hein, J. Eisert, and H. J. Briegel. Multiparty entanglement in graph states. *Phys. Rev. A*, 69:062311, Jun 2004. DOI: 10.1103/PhysRevA.69.062311.

[44] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1):253–258, 1925.

[45] Theodoros Kapourniotis, Elham Kashefi, and Animesh Datta. Blindness and verification of quantum computation with one pure qubit. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 27. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014. DOI: 10.4230/LIPIcs.TQC.2014.176.

[46] Theodoros Kapourniotis, Vedran Dunjko, and Elham Kashefi. On optimising quantum communication in verifiable quantum computing, 2015. In *Proceedings of the 15th Asian Quantum Information Science Conference (AQISC 2015), Seoul, Korea*, pages 25–28, 2015. URL `https://drive.google.com/file/d/0BxR7G6Hj7VPtZ2Y4cHQ1aDcxRlU/view`.

[47] Elham Kashefi and Petros Wallden. Optimised resource construction for verifiable quantum computation. *Journal of Physics A: Mathematical and Theoretical*, 50(14):145306, 2017. DOI: 10.1088/1751-8121/aa5dac.

[48] J Kelly, R Barends, A G Fowler, A Megrant, E Jeffrey, T C White, D Sank, J Y Mutus, B Campbell, Yu Chen, Z Chen, B Chiaro, A Dunsworth, I.-C. Hoi, C Neill, et al. State preservation by repetitive error detection in a superconducting quantum circuit. *Nature*, 519(7541):66–69, mar 2015. ISSN 0028-0836. DOI: 10.1038/nature14270.

[49] T. Lanting, A. J. Przybysz, A. Yu. Smirnov, F. M. Spedalieri, M. H. Amin, A. J. Berkley, R. Harris, F. Altomare, S. Boixo, P. Bunyk, N. Dickson, C. Enderud, J. P. Hilton, E. Hoskinson, M. W. Johnson, et al. Entanglement in a quantum annealing processor. *Phys. Rev. X*, 4:021041, May 2014. DOI: 10.1103/PhysRevX.4.021041.

[50] Ludovico Latmiral, Nicolò Spagnolo, and Fabio Sciarrino. Towards quantum supremacy with lossy scattershot boson sampling. *New Journal of Physics*, 18 (11):113008, 2016. DOI: 10.1088/1367-2630/18/11/113008.

[51] T. D. Lee and C. N. Yang. Statistical theory of equations of state and phase transitions. ii. lattice gas and ising model. *Phys. Rev.*,

87:410–419, Aug 1952. DOI: 10.1103/PhysRev.87.410.

[52] U. Mahadev. Classical verification of quantum computations. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 259–267, Oct 2018. DOI: 10.1109/FOCS.2018.00033.

[53] Atul Mantri, Tommaso F Demarie, and Joseph F Fitzsimons. Universality of quantum computation with cluster states and (x, y)-plane measurements. *Scientific reports*, 7: 42861, 2017. DOI: 10.1038/srep42861.

[54] Damian Markham and Alexandra Krause. A simple protocol for certifying graph states and applications in quantum networks. *arXiv preprint arXiv:1801.05057*, 2018. URL https://arxiv.org/abs/1801.05057.

[55] Rawad Mezher, Joe Ghalbouni, Joseph Dgheim, and Damian Markham. Efficient quantum pseudorandomness with simple graph states. *Phys. Rev. A*, 97:022333, Feb 2018. DOI: 10.1103/PhysRevA.97.022333.

[56] Jacob Miller, Stephen Sanders, and Akimasa Miyake. Quantum supremacy in constant-time measurement-based computation: A unified architecture for sampling and verification. *Physical Review A*, 96(6):062320, 2017. DOI: 10.1103/PhysRevA.96.062320.

[57] Daniel Mills, Anna Pappa, Theodoros Kapourniotis, and Elham Kashefi. Information theoretically secure hypothesis test for temporally unstructured quantum computation. *QPL/IQSA*, 2017. URL https://qpl.science.ru.nl/papers/QPL_2017_paper_42.pdf.

[58] Tomoyuki Morimae. Hardness of classically sampling the one-clean-qubit model with constant total variation distance error. *Physical Review A*, 96(4):040302, 2017. DOI: 10.1103/PhysRevA.96.040302.

[59] Tomoyuki Morimae and Takeshi Koshiba. Impossibility of secure cloud quantum computing for classical client. *Quantum Information and Computation 19, 0214-0221*, 2019. DOI: 10.26421/QIC19.3-4.

[60] Sam Pallister, Noah Linden, and Ashley Montanaro. Optimal verification of entangled states with local measurements. *Physical review letters*, 120(17):170502, 2018. DOI: 10.1103/PhysRevLett.120.170502.

[61] Borja Peropadre, Alan Aspuru-Guzik, and Juan Jose Garcia-Ripoll. Spin models and boson sampling. 2015. URL https://arxiv.org/abs/1509.02703.

[62] John Preskill. Quantum computing and the entanglement frontier. 2012. URL https://arxiv.org/abs/1203.5813.

[63] R Raussendorf, J Harrington, and K Goyal. Topological fault-tolerance in cluster state quantum computation. *New Journal of Physics*, 9(6):199, 2007. DOI: 10.1088/1367-2630/9/6/199.

[64] Robert Raussendorf and Hans J. Briegel. A one-way quantum computer. *Phys. Rev. Lett.*, 86:5188–5191, May 2001. DOI: 10.1103/PhysRevLett.86.5188.

[65] Ben W Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496(7446):456–460, 2013. DOI: 10.1038/nature12035.

[66] Lucile Savary and Leon Balents. Quantum spin liquids: a review. *Reports on Progress in Physics*, 80(1):016502, 2017. DOI: 10.1088/0034-4885/80/1/016502.

[67] Dan Shepherd and Michael J Bremner. Temporally unstructured quantum computation. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 465(2105):1413–1439, 2009. ISSN 1364-5021. DOI: 10.1098/rspa.2008.0443.

[68] Justin B Spring, Benjamin J Metcalf, Peter C Humphreys, W Steven Kolthammer, Xian-Min Jin, Marco Barbieri, Animesh Datta, Nicholas Thomas-Peter, Nathan K Langford, Dmytro Kundys, James C Gates, Brian J Smith, Peter G R Smith, and Ian A Walmsley. Boson sampling on a photonic chip. *Science (New York, N.Y.)*, 339(6121): 798–801, feb 2013. ISSN 1095-9203. DOI: 10.1126/science.1231692.

[69] Larry Stockmeyer. On Approximation Algorithms for # P. *SIAM Journal on Computing*, 14(4):849–861, nov 1985. ISSN 0097-5397. DOI: 10.1137/0214060.

[70] B.M. Terhal and D.V. DiVincenzo. Adaptive quantum computation, constant depth quantum circuits and arthur-merlin games. *Quantum Information and Computation*, 4 (2):134 – 45, 2004. ISSN 1533-7146. DOI: 10.26421/QIC4.2.

A Bridge and break operators and the Ising Sampler

To understand the procedure of carving a specific graph out of a square lattice state by using only xy-plane measurements, we explain two types of operations originally introduced in [29, 43].

The first is *break* operators. Let i be a vertex we want to remove from the original graph, together with the connection to its neighbours. One can achieve this by performing a Pauli Z measurement on the qubit that corresponds to i and discard the outcome. However, we do not have the power to perform measurements out of the xy-plane in our protocol, otherwise we risk revealing the position of the traps by asking the prover to change the basis. Pauli Z-measurements can be simulated by preparing *dummy* qubits in the $|0\rangle$ state. Then,the cZ gate applied by the prover has no effect in entangling it with its neighbours. The measurement can have any arbitrary rotation since the qubit is isolated and does not participate in the computation. In order to keep the whole procedure blind we instead prepare the qubit in state $|d_i\rangle$ for d_i chosen independently and uniformly at random from $\{0,1\}$. This ensures that the state is identical to the maximally mixed state, as is the case for the other qubits in a blind protocol. Also, we apply on each of its neighbours Pauli operation Z^{d_i}, before sending them to the prover, so that we cancel the effect that the prover's entangling will have on that neighbour.

The second is called *bridge* operators. Let i be a vertex of degree 2 that we want to remove in the original graph and join its neighbours by an new edge. To achieve this we apply a Pauli Y measurement (measurement angle $\pi/2$) on the qubit that corresponds to vertex i and add $\pi/2$ on the angles of each the two neighbours. Conditioned on this measurement giving 0 the resulting graph, when we trace out the measured qubit, is the desired one. If the measurement outcome is 1 then, in order to get the correct graph, we need to apply a Z correction on the neighbours. Since our Ising sampler model is nonadaptive and all our measurements are in the xy plane we can achieve this by flipping the measurement outcomes of the corresponding qubits. Notice that this is not an issue in the trap rounds that we explain in Section 4.1 since there are no bridge operations in this case.

B Proof of soundness in Theorem 1

Proof. Let $U_P(\boldsymbol{r},\boldsymbol{d})$ denote the correct unitary operation of the protocol. It includes everything preceding U_B in Fig. (6), and we have only included in the arguments the random parameters that will be averaged over later. The vector $\boldsymbol{r}$ contains as elements bits r_i, r'_i, where i ranges from 1 to $(2\kappa+1)N$ ($2^{2(2\kappa+1)N}$ possible values) and the vector $\boldsymbol{d}$ contains as elements bits d_i, where i ranges again from 1 to $(2\kappa+1)N$ (for the non-dummy qubits we assume fixed $d_i=0$, thus $2^{\kappa N}$ possible values). The rest of the random parameters are the vector $\boldsymbol{\theta}$ which contains elements $\theta_i \in \{0, \frac{\pi}{8}, \frac{2\pi}{8}, \ldots, \frac{15\pi}{8}\}$ for $1 \leq i \leq (2\kappa+1)N$ ($16^{(2\kappa+1)N}$ possible values) and the vector $\boldsymbol{t}$ which contains the indices of the positions of the traps in all 2κ trap systems and takes all $\binom{2\kappa+1}{\kappa}(\kappa+1)$ possible values allowed by the construction. The distributions over all the possible values of the above random parameters are uniform.

In the honest case, after $U_P(\boldsymbol{r},\boldsymbol{d})$ is applied, the state of the trap system becomes $\rho_t = \bigotimes_{i\in\boldsymbol{t}} Z^{r_i} |+\rangle_i \langle+|_i Z^{r_i}$, where the index i takes values from the elements of $\boldsymbol{t}$ that represent the positions of the traps. In the dishonest case (again based on Fig. (6)), tracing out the prover's private system, the deviation U_B becomes an arbitrary CPTP map denoted by $\mathcal{E}$. The probability of getting all zeros of the trap system ρ'_t, right after the measurements, can be written as

$$F_t^2 \equiv \sum_{\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{t},\boldsymbol{d}} p(\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{t},\boldsymbol{d}) \mathrm{Tr}\left(\bigotimes_{i\in\boldsymbol{t}} Z^{r_i} |+\rangle_i \langle+|_i Z^{r_i} \rho'_t\right) \tag{32}$$

$$= \sum_{\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{t},\boldsymbol{d},\boldsymbol{b}} p\mathrm{Tr}\left(\bigotimes_{i\in\boldsymbol{t}} Z^{r_i}\left|+\right\rangle_i\left\langle+\right|_i Z^{r_i}\mathrm{Tr}_{\{i:i\notin\boldsymbol{t}\}}\left(\bigotimes_i Z^{b_i}\left|+\right\rangle_i\left\langle+\right|_i Z^{b_i}\mathcal{E}\left(U_P(\boldsymbol{r},\boldsymbol{d})\bigotimes_{i\notin\boldsymbol{m}(\boldsymbol{t})}\left|+\right\rangle_i\left\langle+\right|_i\right.\right.\right.$$
$$\left.\left.\left.\bigotimes_{i\in\boldsymbol{m}(\boldsymbol{t})}\left|0\right\rangle_i\left\langle0\right|_i U_P(\boldsymbol{r},\boldsymbol{d})^\dagger\bigotimes_i\left|\delta_i(\theta_i,r_i)\right\rangle\left\langle\delta_i(\theta_i,r_i)\right|\right)\bigotimes_i Z^{b_i}\left|+\right\rangle_i\left\langle+\right|_i Z^{b_i}\right)\right) \tag{33}$$

where the inner trace in the formula is taken over all the systems except the trap system. The vector $\boldsymbol{b}$ has been introduced, where elements b_i are bits which correspond to the results of measurements of bits i for $1 \leq i \leq (2\kappa+1)N$. The probability p comes from the uniform distribution over all possible values of the random parameters and is therefore $1/(2^{2(2\kappa+1)N}2^{\kappa N}16^{(2\kappa+1)N}\binom{2\kappa+1}{\kappa}(\kappa+1))$. Also, $\boldsymbol{m}(\boldsymbol{t})$ are the positions of the dummy qubits for a choice of trap positions $\boldsymbol{t}$.

Summing over θ's creates the maximally mixed state for the δ's and summing over the r's and d's of the computational system and the dummy system creates the maximally mixed state for those systems. This is because just before the application of deviation operator these systems are not entangled with the trap system and at the same time a quantum one-time-pad is applied on them. We can therefore trace them out and update the CPTP map $\mathcal{E}$ to a new CPTP map $\mathcal{E}'$ that applies on the remaining system (of dimension $2^{N'}$) and does not depend on the secret parameters.

The CPTP map $\mathcal{E}'$ can be written as a Kraus decomposition, where the Kraus operators $\{E_u\}$ obey $\sum_u E_u E_u^\dagger = I_{2^{N'}}$, where $I_{2^{N'}}$ is the identity on a $2^{N'}$ dimension system. Each Kraus operator can be further decomposed into the Pauli basis as $E_u = \sum_k a_{u,k}P_k$, where $\{P_k\}$ are all generalized elements of the Pauli basis applying on a $2^{N'}$ dimension system and $\{a_{u,k}\}$ are complex coefficients. Also, we remind the reader that the ϕ parameters of the trap qubits are all zero and therefore the remaining honest operation consists only of the rotations by the r parameters.

$$F_t^2 = \sum_{\boldsymbol{r_t},\boldsymbol{t},\boldsymbol{b_t}} p(\boldsymbol{r_t},\boldsymbol{t})\mathrm{Tr}\Bigg(\sum_u\sum_{k=1}^{4^{N'}}\sum_{l=1}^{4^{N'}} a_{u,k}a_{u,l}^*\bigotimes_{i\in\boldsymbol{t}} Z^{r_i}\left|+\right\rangle_i\left\langle+\right|_i Z^{r_i}\bigotimes_{i\in\boldsymbol{t}} Z^{b_i}\left|+\right\rangle\left\langle+\right| Z^{b_i}P_{k|i}Z^{r_i}\left|+\right\rangle_i\left\langle+\right|_i Z^{r_i}P_{l|i}$$
$$Z^{b_i}\left|+\right\rangle\left\langle+\right| Z^{b_i}\Bigg), \tag{34}$$

where $P_{k|i}$ denotes the Pauli operator that applies on qubit with index i from the generalized Pauli basis operator P_k. Because of the state of the system, in particular the fact that $X\left|+\right\rangle = \left|+\right\rangle$, we can add 'free' Pauli X operators randomized by new parameters r' taken uniform at random.

$$F_t^2 = \sum_{\boldsymbol{r_t},\boldsymbol{r'_t},\boldsymbol{t},\boldsymbol{b_t}} p(\boldsymbol{r_t},\boldsymbol{r'_t},\boldsymbol{t})\mathrm{Tr}\Bigg(\sum_{u,k,l} a_{u,k}a_{u,l}^*\bigotimes_{i\in\boldsymbol{t}} Z^{r_i}\left|+\right\rangle_i\left\langle+\right|_i Z^{r_i}\bigotimes_{i\in\boldsymbol{t}} Z^{b_i}\left|+\right\rangle\left\langle+\right| X^{r'_i}Z^{b_i}P_{k|i}Z^{r_i}X^{r'_i}\left|+\right\rangle_i\left\langle+\right|_i$$
$$X^{r'_i}Z^{r_i}P_{l|i}Z^{b_i}X^{r'_i}\left|+\right\rangle\left\langle+\right| Z^{b_i}\Bigg) \tag{35}$$

By changing variables $b'_i = b_i + r_i$ and applying the cyclic property of the trace to move $Z^{r_i}X^{r'_i}$ around

$$F_t^2 = \sum_{\boldsymbol{r_t},\boldsymbol{r'_t},\boldsymbol{t},\boldsymbol{b'_t}} p(\boldsymbol{r_t},\boldsymbol{r'_t},\boldsymbol{t})\mathrm{Tr}\Bigg(\sum_{u,k,l} a_{u,k}a_{u,l}^*\bigotimes_{i\in\boldsymbol{t}}\left|+\right\rangle_i\left\langle+\right|_i\bigotimes_{i\in\boldsymbol{t}} Z^{b'_i}\left|+\right\rangle\left\langle+\right| X^{r'_i}Z^{b'_i+r_i}P_{k|i}Z^{r_i}X^{r'_i}\left|+\right\rangle_i\left\langle+\right|_i$$
$$X^{r'_i}Z^{r_i}P_{l|i}Z^{b'_i+r_i}X^{r'_i}\left|+\right\rangle\left\langle+\right| Z^{b'_i}\Bigg) \tag{36}$$

Applying the Pauli twirl lemma [21], proven in Appendix E, by averaging over $\boldsymbol{r_t},\boldsymbol{r'_t}$, we get

$$F_t^2 = \sum_{\boldsymbol{t},\boldsymbol{b}'_{\boldsymbol{t}}} p(\boldsymbol{t}) \sum_{u,k} |a_{u,k}|^2 \prod_{i\in\boldsymbol{t}} |\langle +|_i Z^{b'_i} |+\rangle \langle +| Z^{b'_i} P_{k|i} |+\rangle_i|^2$$
$$= \sum_{\boldsymbol{t}} p(\boldsymbol{t}) \sum_{k} |\alpha_k|^2 \prod_{i\in\boldsymbol{t}} |\langle +|_i P_{k|i} |+\rangle_i|^2, \tag{37}$$

where $|\alpha_k|^2 = \sum_u |a_{u,k}|^2$ and $\sum_k |\alpha_k|^2 = 1$ from the unital property of the attack.

A similar analysis is applied to calculate the average fidelity $F_c = F(\rho_c, \rho'_c)$ of the computational state after the measurements. In the honest case the computational state ρ_c just before the measurement will be disentangled from the rest of the system: $\bigotimes_{i\in\boldsymbol{c}} Z^{r_i} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle\langle G| \bigotimes_{j\in\boldsymbol{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} Z^{r_j}$, where $\boldsymbol{c(t)}$ are the positions of the computational qubits for a choice of trap positions $\boldsymbol{t}$ and $|G\rangle\langle G|$ is the computational graph state. The latter can be expressed as $E_G \bigotimes_{i\in\boldsymbol{c}} |+\rangle_i \langle +|_i E_G^\dagger$, where E_G denotes all entangling operators cZ that apply on a graph G. In the dishonest case, for an attack $\mathcal{E}$ the fidelity $\bar{F}_c$ is

$$F_c \equiv \sum_{\boldsymbol{r},\boldsymbol{r'},\boldsymbol{\theta},\boldsymbol{t},\boldsymbol{d}} p(\boldsymbol{r},\boldsymbol{r'},\boldsymbol{\theta},\boldsymbol{t},\boldsymbol{d}) F(\rho_c,\rho'_c) \tag{38}$$
$$= \sum_{\boldsymbol{r},\boldsymbol{r'},\boldsymbol{\theta},\boldsymbol{t},\boldsymbol{d}} p(\boldsymbol{r},\boldsymbol{r'},\boldsymbol{\theta},\boldsymbol{t},\boldsymbol{d}) \mathrm{Tr}\left((\rho_c\rho'_c)^{1/2}\right) \tag{39}$$
$$\geq \mathrm{Tr}\left(\left(\sum_{\boldsymbol{r},\boldsymbol{r'},\boldsymbol{\theta},\boldsymbol{t},\boldsymbol{d}} p(\boldsymbol{r},\boldsymbol{r'},\boldsymbol{\theta},\boldsymbol{t},\boldsymbol{d})\rho_c\rho'_c\right)^{1/2}\right) \tag{40}$$
$$= \mathrm{Tr}\left(\left(\sum_{\boldsymbol{r},\boldsymbol{r'},\boldsymbol{\theta},\boldsymbol{t},\boldsymbol{d},\boldsymbol{b}} p \bigotimes_{i\in\boldsymbol{c}} Z^{r_i} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle\langle G| \bigotimes_{j\in\boldsymbol{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} Z^{r_j} \mathrm{Tr}_{\{i:i\notin\boldsymbol{c}\}} \left(\bigotimes_i Z^{b_i} |+\rangle_i \right.\right.\right.$$
$$\left.\left.\left. \langle +|_i Z^{b_i} \mathcal{E}(U_P(\boldsymbol{r},\boldsymbol{r'},\boldsymbol{d}) \bigotimes_{i\notin\boldsymbol{m(t)}} |+\rangle_i \langle +|_i \bigotimes_{i\in\boldsymbol{m(t)}} |0\rangle_i \langle 0|_i U_P(\boldsymbol{r},\boldsymbol{r'},\boldsymbol{d})^\dagger \bigotimes_i |\delta_i(\theta_i,r_i,r'_i)\rangle\langle\delta_i(\theta_i,r_i,r'_i)|) \right.\right.\right.$$
$$\left.\left.\left. \bigotimes_i Z^{b_i} |+\rangle_i \langle +|_i Z^{b_i} \right)\right)^{1/2}\right). \tag{41}$$

Summing over the θ's of the δ's and the r's and the d's of the trap and the dummy system creates the maximally mixed system for these systems which can be traced over. Expressing the attack on the remaining system (of dimension $2^{N'}$) using the Kraus decomposition with each Kraus element decomposed in the Pauli basis we get as before

$$F_c \geq \mathrm{Tr}\Bigg(\Bigg(\sum_{\boldsymbol{t},\boldsymbol{r_{c(t)}},\boldsymbol{r'_{c(t)}},\boldsymbol{b_{c(t)}}} p \sum_u \sum_{k=1}^{4^{N'}} \sum_{l=1}^{4^{N'}} a_{u,k} a^*_{u,l} \bigotimes_{i\in\boldsymbol{c}} Z^{r_i} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle\langle G| \bigotimes_{j\in\boldsymbol{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} Z^{r_j}$$
$$\bigotimes_{i\in\boldsymbol{c}} Z^{b_i} |+\rangle\langle +| Z^{b_i} P_{k|i} Z^{r_i} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle\langle G| \bigotimes_{j\in\boldsymbol{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} Z^{r_j} P_{l|j} Z^{b_i} |+\rangle\langle +| Z^{b_i} \Bigg)^{1/2}\Bigg) \tag{42}$$

By changing variables $b' = b + r$ and applying the cyclic property of the trace

$$F_c \geq \mathrm{Tr}\Bigg(\Bigg(\sum_{\boldsymbol{t},\boldsymbol{r_{c(t)}},\boldsymbol{r'_{c(t)}},\boldsymbol{b'_{c(t)}}} p \sum_u \sum_{k=1}^{4^{N'}} \sum_{l=1}^{4^{N'}} a_{u,k} a^*_{u,l} \bigotimes_{i\in\boldsymbol{c}} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle\langle G| \bigotimes_{j\in\boldsymbol{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} \bigotimes_{i\in\boldsymbol{c}} Z^{b'_i}$$
$$|+\rangle\langle +| Z^{b'_i+r_i} P_{k|i} Z^{r_i} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle\langle G| \bigotimes_{j\in\boldsymbol{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} Z^{r_j} P_{l|j} Z^{b'_i+r_i} |+\rangle\langle +| Z^{b'_i} \Bigg)^{1/2}\Bigg) \tag{43}$$

Using Corollary 1 in Appendix E and the cyclic property of the trace and sum over $\boldsymbol{r_{c(t)}}$, we can eliminate all Pauli X operators of the attack that differ in the two sides (we denote this by replacing the summation over l with a summation over l_x where the element P_l agrees with P_k on all the Pauli X components). We also use the cyclic property of the trace to get

$$F_c \geq \mathrm{Tr}\Bigg(\Bigg(\sum_{\boldsymbol{t},\boldsymbol{r_{c(t)}},\boldsymbol{r'_{c(t)}},\boldsymbol{b'_{c(t)}}} p\sum_u\sum_k\sum_{l_x} a_{u,k}a^*_{u,l}\,\langle G|\bigotimes_{j\in\boldsymbol{c}} X^{r'_j}R_z(-\phi_j)X^{r'_j}\bigotimes_{i\in\boldsymbol{c}} Z^{b'_i}\,|+\rangle\langle+|\,Z^{b'_i}P_{k|i}X^{r'_i}R_z(\phi_i)$$
$$X^{r'_i}\,|G\rangle\langle G|\bigotimes_{j\in\boldsymbol{c}} X^{r'_j}R_z(-\phi_j)X^{r'_j}P_{l|j}Z^{b'_i}\,|+\rangle\langle+|\,Z^{b'_i}\bigotimes_{i\in\boldsymbol{c}} X^{r'_i}R_z(\phi_i)X^{r'_i}\,|G\rangle\Bigg)^{1/2}\Bigg). \tag{44}$$

Then, the $X^{r'_i}$ ($X^{r'_j}$) operators that are next to $|G\rangle$ ($\langle G|$) can be rewritten as Pauli Z operators on their neighbours, which then commute with z rotations and the attack (since the Pauli X operators of the attack are the same from both sides) and with the projectors $Z^{b'_i}\,|+\rangle\langle+|\,Z^{b'_i}$, by changing variable $b''_i = b'_i + r'_i$, and cancel each other. The $X^{r'_i}$ operators that are next to projectors $Z^{b'_i}\,|+\rangle\langle+|\,Z^{b'_i}$ commute with them trivially

$$F_c \geq \mathrm{Tr}\Bigg(\Bigg(\sum_{\boldsymbol{t},\boldsymbol{r'_{c(t)}},\boldsymbol{b''_{c(t)}}} p\sum_u\sum_k\sum_{l_x} a_{u,k}a^*_{u,l}\,\langle G|\bigotimes_{j\in\boldsymbol{c}} R_z(-\phi_j)\bigotimes_{i\in\boldsymbol{c}} Z^{b''_i}\,|+\rangle\langle+|\,Z^{b''_i}$$
$$X^{r'_j}P_{k|i}X^{r'_i}R_z(\phi_i)\,|G\rangle\langle G|\bigotimes_{j\in\boldsymbol{c}} R_z(-\phi_j)X^{r'_j}P_{l|j}X^{r'_i}Z^{b''_i}\,|+\rangle\langle+|\,Z^{b''_i}\bigotimes_{i\in\boldsymbol{c}} R_z(\phi_i)\,|G\rangle\Bigg)^{1/2}\Bigg). \tag{45}$$

Then we can use again Corollary 1, but with Q, Q' being Pauli Z+identity operators and $\{P_i\}$ all tensor products of Pauli X+identity operators, and sum over $\boldsymbol{r'_{c(t)}}$ to eliminate the Pauli Z components of the attack the differ in the two sides. Thus, given that $\sum_{u,k}|a_{u,k}|^2 = 1$ from the unital property of the attack, the attack becomes a convex combination of Pauli operators:

$$F_c \geq \Bigg(\sum_{\boldsymbol{t},\boldsymbol{b''_{c(t)}}} p(\boldsymbol{t})\sum_{u,k}|a_{u,k}|^2\bigotimes_{i\in\boldsymbol{c(t)}}\langle+|_i\,E_G^\dagger\bigotimes_{i\in\boldsymbol{c(t)}} R_z(-\phi_i)Z^{b''_i}\,|+\rangle\langle+|\,Z^{b''_i}P_{k|i}R_z(\phi_i)E_G\bigotimes_{i\in\boldsymbol{c(t)}}|+\rangle_i\,|^2\Bigg)^{1/2} \tag{46}$$

The Pauli X component of $P_{k|i}$ can be replaced by I since the only effect is, depending on b''_i, to change the sign of the quantity inside the absolute and the sign is eliminated. Then, we can sum over the b's to get identity and since $E_G^\dagger R_z(-\phi_i)$ now commutes with the attack Pauli operators, it cancels with $R_z(\phi_i)E_G$. Also, as before, we set $|\alpha_k|^2 = \sum_u |a_{u,k}|^2$

$$F_c^2 \geq \sum_{\boldsymbol{t}} p(\boldsymbol{t})\sum_k |\alpha_k|^2\prod_{i\in\boldsymbol{c(t)}}|\,\langle+|_i\,P_{k|i}\,|+\rangle_i\,|^2. \tag{47}$$

It is easy to see by the symmetry of the trap construction that, for attacks that are exactly the same on any of the $2\kappa+1$ graphs of the different the rounds of the protocol e.g. stochastic noise, Equation 37 and Equation 47 give the same result when averaged over $\boldsymbol{t}$. However, one needs to deal with more clever attacks which attack a different qubit at every round trying to coincide with dummies instead of traps with non-zero probability, i.e. for some of the possible permutations of the $2\kappa+1$ graphs.

We now show that for $\kappa \geq 1$ the maximum of $F_t^2 - F_c^2$ for all possible deviation strategies is

$$\Delta_\kappa \equiv \frac{\kappa!(\kappa+1)!}{(2\kappa+1)!}$$

The attack is a convex combination of Pauli operators, thus it suffices to find the Pauli operator that maximizes $F_t^2 - F_c^2$. The maximum comes from the attack that touches all $2\kappa+1$ rounds. In this case, F_c^2 is lower bounded by 0 and F_t^2 comes from the probability the attack does not coincide with any trap in the 2κ trap computation rounds. There are $2\kappa+1$ ways of picking the target computation round, which fixes the positions of the 2κ trap computation rounds. The choice of the even/odd parity positions for the traps is fixed not to coincide with the attacks. Further simplification can be done

by observing that only the attacks on $\kappa+1$ of the same kind (even or odd) and κ of the other, are successful. By attacking $\kappa+2$ or more of the same kind they are guaranteed to hit a trap of the same kind, independently of the position of the target. This reduces the possible ways of picking the target to $\kappa+1$, which gives a bound of $\frac{\kappa+1}{\binom{2\kappa+1}{\kappa}(\kappa+1)}$, equal to the value of Δ_κ above.

To conclude the proof we show that for all other attacks $F_t^2 - F_c^2$ is non-positive and thus estimating F_c^2 through F_t^2 is a conservative estimation. We begin by arguing that there is no benefit for the attacker to touch more than one qubit of each round, since the lower bound of the fidelity F_c^2 contains products of terms that can be 0 or 1 and thus it suffices to make one term 0 to make the product 0. By symmetry of the construction it does not matter which particular qubit the attacker touches but only whether it is an odd or even position qubit at each round. Let λ be the number of rounds that are attacked.

Assume, without loss of generality, the first ξ rounds are attacked on an even qubit, where $\xi \leq \kappa$ otherwise it will certainly hit a trap, and $\lambda - \xi \leq \kappa$ for the same reason. Also, assume $\xi \geq (\lambda - \xi)$ without loss of generality.

For index k to correspond to an attack on λ rounds

$$F_{c,k}^2 \geq 1 - \frac{\lambda}{2\kappa+1} = \frac{2\kappa+1-\lambda}{2\kappa+1} \tag{48}$$

There are $\binom{2\kappa+1}{\kappa}(\kappa+1)$ possibilities for the selection of traps. In order to count the combinations of attacks not affecting the traps, we identify two cases, (i) the target is in the attacked rounds (λ possible positions) and there are $\binom{2\kappa+1-\lambda}{\kappa-\xi}$ possible placings of the remaining traps in the non-attacked positions, (ii) the target is not in the attacked rounds ($2\kappa+1-\lambda$ possible positions) and there are $\binom{2\kappa-\lambda}{\kappa-\xi}$ possible placings of the remaining traps in the non-attacked positions:

$$F_{t,k} = \frac{\lambda\binom{2\kappa+1-\lambda}{\kappa-\xi} + (2\kappa+1-\lambda)\binom{2\kappa-\lambda}{\kappa-\xi}}{\binom{2\kappa+1}{\kappa}(\kappa+1)} \tag{49}$$

We show that for $\lambda \leq 2\kappa$ we have $F_{t,k} - F_{c,k} \leq 0$.

$$\begin{aligned}
\frac{\lambda\binom{2\kappa+1-\lambda}{\kappa-\xi} + (2\kappa+1-\lambda)\binom{2\kappa-\lambda}{\kappa-\xi}}{\binom{2\kappa+1}{\kappa}(\kappa+1)} - \frac{2\kappa+1-\lambda}{2\kappa+1} \leq 0 \Leftrightarrow \\
\frac{(2\kappa-\lambda+1)(\kappa+\xi+1)\binom{2\kappa-\lambda}{\kappa-\xi}}{(\kappa-(\lambda-\xi)+1)\binom{2\kappa+1}{\kappa}(\kappa+1)} - \frac{2\kappa+1-\lambda}{2\kappa+1} \leq 0 \Leftrightarrow \\
\frac{(2\kappa-\lambda+1)(\kappa+\xi+1)(2\kappa-\lambda)!(\kappa!)^2}{(\kappa-(\lambda-\xi)+1)(2\kappa+1)!(\kappa-(\lambda-\xi))!(\kappa-\xi)!} \leq \frac{2\kappa+1-\lambda}{2\kappa+1} \Leftrightarrow \\
\frac{(\kappa+\xi+1)(2\kappa-\lambda)!(\kappa!)^2}{(\kappa-(\lambda-\xi)+1)(2\kappa)!(\kappa-(\lambda-\xi))!(\kappa-\xi)!} \leq 1
\end{aligned} \tag{50}$$

For $\kappa = \{1, 2\}$ it is easy to verify the expression directly. For the general case we rewrite the above as:

$$\frac{(\kappa+\xi+1)[(\kappa-\xi+1)\cdot\ldots\cdot\kappa][(\kappa-(\lambda-\xi)+1)\cdot\ldots\cdot\kappa]}{(\kappa-(\lambda-\xi)+1)(2\kappa-\lambda+1)\cdot\ldots\cdot 2\kappa} \leq 1 \tag{51}$$

where we have $1+\xi+(\lambda-\xi) = 1+\lambda$ terms on the numerator and $1+\lambda$ terms in the denominator. For the LHS of the above equation we have

$$\leq \frac{(\kappa+\xi+1)}{(\kappa-(\lambda-\xi)+1)2^{\xi}}$$
$$\leq \frac{(\kappa+\xi+1)}{(\kappa-\xi+1)2^{\xi}} \tag{52}$$

It suffices to show that the above is ≤ 1 for all allowed values of κ and ξ. We can rewrite it as:

$$\frac{(\kappa+\xi+1)}{(\kappa-\xi+1)} \leq 2^{\xi} \Leftrightarrow \tag{53}$$
$$\frac{1}{\ln(2)} \ln(\frac{\kappa+1+\xi}{\kappa+1-\xi}) \leq \xi \Leftrightarrow \tag{54}$$
$$\frac{2}{\ln(2)} \text{artanh}\left(\frac{\xi}{\kappa+1}\right) \leq \xi \tag{55}$$

which is true for $\kappa \geq 3, \xi \leq \kappa$. This concludes our calculation of the bound Δ_{κ}.
The rest of the proof is given in the main text.

□

C Calculation of FT Threshold for Protocol 2b

Since the logical graph is the brickwork state, its topological implementation will be on MBQC, as shown in Fig. (8), and will have the structure of Fig. (11).

Noise considered is local, unital and bounded. It applies after every elementary operation (preparation, entangling and measurement) j and is expressed as a CPTP superoperator:

$$\mathcal{N}_j = (1-\epsilon)\mathcal{I} + \mathcal{E}_j \tag{56}$$

where $||\mathcal{E}_j||_{\diamond} = \epsilon$, where we set $\epsilon = \epsilon_V = \epsilon_P$ to calculate a common threshold for the verifier and the prover.

For the fault tolerant noisy, but honest, probability distribution post-selected for null syndrome measurement $q^{\text{nsy}}(\boldsymbol{x}|y=0)$, and the exact one $q^{\text{exc}}(\boldsymbol{x})$ we reproduce the derivation of Ref. [32].

$$\sum_{\boldsymbol{x}} |q^{\text{nsy}}(\boldsymbol{x}|y=0) - q^{\text{exc}}(\boldsymbol{x})| = \sum_{\boldsymbol{x}} \left| \frac{\text{Tr}(P_{\boldsymbol{x}} Q_y(\rho_{\text{sparse}} + \rho_{\text{faulty}}))}{\text{Tr}(Q_y(\rho_{\text{sparse}} + \rho_{\text{faulty}}))} - q^{\text{exc}}(\boldsymbol{x}) \right|$$
$$= \sum_{\boldsymbol{x}} \left| \frac{\text{Tr}(P_{\boldsymbol{x}} Q_y \rho_{\text{faulty}})}{\text{Tr}(Q_y(\rho_{\text{sparse}} + \rho_{\text{faulty}}))} - (1-b) q^{\text{exc}}(\boldsymbol{x}) \right| \tag{57}$$

Here, $P_{\boldsymbol{x}}$ is the projector to result $\boldsymbol{x}$ for the output register and Q_y is the projector to null syndrome for the post-selection register. The output quantum state of the sampler, just before the final measurements, can be written as a sum of two matrices: ρ_{sparse} that is the sum of the states on which apply the components of the noise operators $\mathcal{N}_j$ (i.e. either component $(1-\epsilon)\mathcal{I}$ or component $\mathcal{E}_j$ for each j) that, under postelection for $y=0$, do not produce a logical error in the output distribution, and ρ_{faulty} which contains the sum of the states on which apply the rest of the noise components (for more detail see [32]). Thus, $\text{Tr}(P_{\boldsymbol{x}} Q_y \rho_{\text{sparse}}) \propto q^{\text{exc}}(\boldsymbol{x})$. Term b, which is defined by:

$$b = \frac{\text{Tr}(P_{\boldsymbol{x}} Q_y \rho_{\text{sparse}})}{\text{Tr}(Q_y(\rho_{\text{sparse}} + \rho_{\text{faulty}})) q^{\text{exc}}(\boldsymbol{x})} \tag{58}$$

is therefore independent of $\boldsymbol{x}$.

By applying triangle inequality and by observing that the trace terms are positive

$$\sum_{\boldsymbol{x}} |q^{\text{nsy}}(\boldsymbol{x}|y=0) - q^{\text{exc}}(\boldsymbol{x})| \leq \frac{\text{Tr}(Q_y \rho_{\text{faulty}})}{\text{Tr}(Q_y(\rho_{\text{sparse}} + \rho_{\text{faulty}}))} + (1-b)$$

Since $\sum_x \frac{\text{Tr}(P_x Q_y(\rho_{\text{sparse}}+\rho_{\text{faulty}}))}{\text{Tr}(Q_y(\rho_{\text{sparse}}+\rho_{\text{faulty}}))} = 1$, we have $1-b = \frac{\text{Tr}(Q_y \rho_{\text{faulty}})}{\text{Tr}(Q_y(\rho_{\text{sparse}}+\rho_{\text{faulty}}))}$.

Also we have $\text{Tr}(Q_y(\rho_{\text{sparse}} + \rho_{\text{faulty}})) \geq (1-\epsilon)^N$. This comes from the fact there is at least one selection of components of the noise operators $\mathcal{N}_j$ that results in the null syndrome and this is components $(1-\epsilon)\mathcal{I}, \forall j$, giving a term with trace $(1-\epsilon)^N$. Thus

$$\sum_{\boldsymbol{x}} |q^{\text{nsy}}(\boldsymbol{x}|y=0) - q^{\text{exc}}(\boldsymbol{x})| \leq 2\text{Tr}(\rho_{\text{faulty}})/(1-\epsilon)^N \tag{59}$$

In the case of the topological code the errors are created by error chains $\mathcal{L}$ of length greater than L_d, which is the minimum of the distance between two defects and the size of defects. Matrix ρ_{faulty} can be decomposed into terms $\rho^{\mathcal{L}}_{\text{faulty}}$ with respect to error chains $\mathcal{L}$ of length L so that:

$$\text{Tr}(\rho_{\text{faulty}}) \leq \sum_{L=L_d}^{N} \sum_{\mathcal{L}:|\mathcal{L}|=L} \text{Tr}(\rho^{\mathcal{L}}_{\text{faulty}}) \tag{60}$$

$$\leq \sum_{L=L_d}^{N} \sum_{\mathcal{L}:|\mathcal{L}|=L} (1-\epsilon)^N \prod_{j=1}^{|\mathcal{L}|} \frac{||\mathcal{E}_j||_\diamond}{1-\epsilon} \tag{61}$$

$$= \sum_{L=L_d}^{N} \sum_{\mathcal{L}:|\mathcal{L}|=L} (1-\epsilon)^N \left(\frac{\epsilon}{1-\epsilon}\right)^{|\mathcal{L}|} \tag{62}$$

The number of error chains of length $|\mathcal{L}|$ in the 3D lattice of size n is $\text{poly}(n)(6/5)5^{|\mathcal{L}|}$, which is the number of self avoiding walks [22]. Thus the gap $\sum_{\boldsymbol{x}} |q^{\text{nsy}}(\boldsymbol{x}|y=0) - q^{\text{exc}}(\boldsymbol{x})|$ is bounded by

$$\leq 2\sum_{L=d}^{N} \text{poly}(N)(6/5)5^L \left(\frac{\epsilon}{1-\epsilon}\right)^L \tag{63}$$

which converges to zero if $\epsilon/(1-\epsilon) < 1/5$. The threshold comes from the self-avoiding walks that affect the singular qubits and surpass the distillation threshold, where a more careful counting needs to be done [32] to get $\epsilon/(1-\epsilon) < 0.134$ or $\epsilon < 0.118$. This calculation of the threshold [32] is an underestimate because of the assumption that error correction is done on primal and dual cubic lattices independently.

However, the above calculation is under the stochastic phenomenological noise model that does not account for the noise in the individual operations that compose a syndrome measurement. In our case, the topological code is implemented on the MBQC model where physical noise should be at least 6 times less than phenomenological noise. This is because there are typically 6 operations involved in a syndrome measurement: 1 syndrome qubit preparation, 4 entangling operations with the surrounding qubits (less on boundaries) and 1 syndrome measurement. This gives us threshold $\epsilon_{\text{thres}} = 0.0196943$.

The overhead of the error detection scheme comes by counting the number of error syndromes that are influencing one trap by catching potentially detectable errors (chains of size $\leq d$). This is the area of dimension d around the defect qubits and their 'past' in terms of MBQC flow (physical layer) which we choose in the smallest of the three dimensions of the topological code. In Fig. (11)(c) we depict the logical qubits (made of prime/dual physical defects) that compose a 'H' shaped component of the brickwork state. In our trappification scheme any topologically protected qubit can be a trap. For a worst case analysis we consider the biggest logical qubit, in terms of number of defects that make it, which is a prime qubit in our example (bottom middle blue loop in Fig. (11)(c)).

The number of syndrome measurements (cubes) will depend on the distance parameter d. Since the counting of syndromes is involved we give an example for fixed values, $d = 2$ and physical noise $\epsilon = (1/20)\epsilon_{\text{thres}}$. In this case the number of syndromes is at most 564 and the number of repetitions is $M = 1/(1-p_c)^{564}$, where p_c is the probability of a cube syndrome failing. Probability p_c is given by $(1-(1-2(6\epsilon))^6)/2$, which is the probability of having an odd number of errors in the 6 faces of a cube. This gives $M \approx 3 \times 10^8$. Overheads for other fractions of the threshold for noise appear in the main text.

D Proof of Theorem 2 soundness

Proof. To establish soundness we need to show that a lower bound in the fidelity on the target computation round and the acceptance probability of the trap computation rounds are the same averaged over the random parameters.

The total variation distance between the experimental (noisy and potentially dishonest) distribution of the Ising sampler $q^{\text{nsy}}(\boldsymbol{x}|y=0)$, where $y=0$ implies conditioning on the null syndrome, and the exact one $q^{\text{exc}}(\boldsymbol{x}|y=0) = q^{\text{exc}}(\boldsymbol{x})$ after the measurements is

$$\text{var}^{\text{Post}} = \frac{1}{2}\sum_{\boldsymbol{x}} |q^{\text{exc}}(\boldsymbol{x}|y=0) - q^{\text{nsy}}(\boldsymbol{x}|y=0)| = \frac{1}{2}\sum_{\boldsymbol{x}} \left| q^{\text{exc}}(\boldsymbol{x}) - \frac{q^{\text{nsy}}(\boldsymbol{x}, y=0)}{q^{\text{nsy}}(y=0)} \right| \tag{64}$$

$$= D\left(\rho_c, \frac{\rho_c^{'\text{post}}}{q^{\text{nsy}}(y=0)}\right) \tag{65}$$

$$\leq \sqrt{1 - F^2\left(\rho_c, \frac{\rho_c^{'\text{post}}}{q^{\text{nsy}}(y=0)}\right)} \tag{66}$$

$$= \sqrt{1 - \text{Tr}^2\left(\sqrt{\frac{\rho_c \rho_c^{'\text{post}}}{q^{\text{nsy}}(y=0)}}\right)}, \tag{67}$$

where ρ_c is the correct state and $\rho_c^{'\text{post}}$ the experimental state, post-selected on the null syndrome measurements, after all measurements. For the rest of this section we denote $q^{\text{nsy}}(y=0)$ as q_0' for simplicity.

For the target round, the average fidelity F_c is calculated in the physical level of the computation as in the non-fault-tolerant case. The qubits are pre-rotated by θ_i, or flipped by d_i in the case of dummies.

Noise can enter either during the state preparation from the verifier, or during the single round elementary MBQC operations (entangling and measurement) of the prover. We assume a noise model which is local, unital and bounded, so that standard fault tolerance techniques are applicable. Noise can always moved after every elementary operation on qubit j and expressed as a CPTP superoperator applies only on the state of qubit j:

$$\mathcal{N}_j = (1-\epsilon)\mathcal{I} + \mathcal{E}_j \tag{68}$$

where $||\mathcal{E}_j||_\diamond \leq \epsilon_{\text{thres}}$.

Crucially, we assume that the noise during the preparation does not have any dependence on the secret parameter θ_i.

Moving all the noise operators just before the measurement, results to a different set of local, unital and bounded operators $\mathcal{N}'_j$, collectively represented as $\mathcal{N}'$.

We apply the same twirling steps as in the proof of Theorem 1 to twirl the CPTP map that is the composition of the attack and the noise. Notice that the twirl on the post-selected qubits is trivial

since there is no sum over b'_i. Thus,

$$F_c^2\left(\rho_c^{\text{post}}, \frac{\rho_c^{'\text{post}}}{q'_0}\right) \geq \frac{1}{q'_0}\sum_{\boldsymbol{t},\boldsymbol{b'_{c(t)}}} p(\boldsymbol{t})\sum_{u,k}|a_{u,k}|^2 \left|\bigotimes_{i\in\boldsymbol{c(t)}}\langle 0|_i \bigotimes_{i\in\boldsymbol{c(t)}}\langle +|_i E_G^\dagger \bigotimes_{i\in\boldsymbol{c(t)}} R_z(-\phi_i)Z^{b'_i}|+\rangle_i \langle +|_i Z^{b'_i}\right.$$
$$\left. P_{k|i}R_z(\phi_i)E_G \bigotimes_{i\in\boldsymbol{c(t)}}|+\rangle_i \bigotimes_{i\in\boldsymbol{c(t)}}|0\rangle_i\right|^2, \tag{69}$$

where b'_i's take fixed values in the sum for the syndrome measurements such that the syndrome indicates null errors.

The only (noise and attack) Pauli operators that have an effect on the above quantity are tensor products of identity and Pauli Z. These operators flip the measurement outcome of the particular qubit. Detectable attacks disappear because of the projector to null syndromes. The undetected attacks that come from operators $P_{k|i}$ can be written as logical bit flips on the subsequent measurements - since it will affect the classical post-processing. Also, the normalization factor vanishes when we trace over the syndrome systems.

Therefore, at the logical level

$$F_c^2 \geq \sum_{\boldsymbol{t},\boldsymbol{b''_{c(t)}}} p(\boldsymbol{t})\sum_{u,k}|a_{u,k}|^2 \left|\bigotimes_{i\in\boldsymbol{c(t)}}\langle 0|_i^L \bigotimes_{i\in\boldsymbol{c(t)}}\langle +|_i^L E_G^{\dagger L} \bigotimes_{i\in\boldsymbol{c(t)}} R_z(-\phi_i^L)Z^{b''_i L}|+\rangle^L \langle +|^L Z^{b''_i L}\right.$$
$$\left. P_{k|i}^L R_z(\phi_i^L)E_G^L \bigotimes_{i\in\boldsymbol{c(t)}}|+\rangle_i^L \bigotimes_{i\in\boldsymbol{c(t)}}|0\rangle_i^L\right|^2, \tag{70}$$

We can now sum over the index $b''_{c(t)}$ to simplify the expression, by cancelling also the rotation and entangling operators. On the logical dummy system the logical Pauli Z attacks have no effect, therefore it has trace 1 and can be simplified to

$$F_c^2 \geq \sum_{\boldsymbol{t}} p(\boldsymbol{t})\sum_{u,k}|a_{u,k}|^2 \prod_{i\in\boldsymbol{c(t)}} |\langle +|_i^L P_{k|i}^L |+\rangle_i^L|^2. \tag{71}$$

The same technique can be employed for the trap rounds, with the difference that instead of post-selection there is error correction for Protocol 2a and error detection for Protocol 2b that results in the same logical state for the same (noise and attack) Pauli operators.

From completeness we have set the limit of acceptance of the fidelity estimate to $(1-2\epsilon'')$. By repeating $N = \log(1/\beta)/(2\epsilon''^2)$ times gets us $\sqrt{\epsilon''}$-close in our estimation with confidence $1-\beta$. Thus, with this confidence we get bound $\text{var}^{\text{Post}} \leq \sqrt{3\epsilon'' + \Delta_\kappa}$.

□

E Channel Twirl Lemma

The following lemma is used in the verifiability proofs.

Lemma 1.

$$\sum_{i=1}^{4^n} P_i Q P_i \rho P_i Q' P_i = 0, \text{ if } Q \neq Q' \tag{72}$$

where ρ is a matrix of dimension $2^n \times 2^n$, Q, Q' are two arbitrary n-fold tensor products of Pauli+identity operators $\{I, X, Y, Z\}$, and $\{P_i\}$ is the set of all n-fold tensor products of Pauli operators and the identity $\{I, X, Y, Z\}$.

A proof of this lemma is also provided in Ref. [21].

Proof. We can write Q as $Z_{\boldsymbol{a}}X_{\boldsymbol{a}'} = Z^{a_1} \otimes \ldots \otimes Z^{a_n}X^{a'_1} \otimes \ldots \otimes X^{a'_n}$, for arbitrary binary vectors $\boldsymbol{a} = (a_1, \ldots, a_n)$, $\boldsymbol{a}' = (a'_1, \ldots, a'_n)$, and similarly $Q' = Z_{\boldsymbol{b}}X_{\boldsymbol{b}'}$. Assuming $Q \neq Q'$, either $\boldsymbol{a} \neq \boldsymbol{b}$ or $\boldsymbol{a}' \neq \boldsymbol{b}'$. Summing over all $P_{\boldsymbol{k},\boldsymbol{k}'}$'s which are the n-fold tensor products of the form $Z_{\boldsymbol{k}}X_{\boldsymbol{k}'} = Z^{k'_1} \otimes \ldots \otimes Z^{k'_n}X^{k_1} \otimes \ldots \otimes X^{k_n}$ for binary vectors $\boldsymbol{k} = (k_1, \ldots, k_n), \boldsymbol{k}' = (k'_1, \ldots, k'_n)$, we get

$$\begin{aligned}
&\sum_{\boldsymbol{k},\boldsymbol{k}'} P_{\boldsymbol{k},\boldsymbol{k}'}QP_{\boldsymbol{k},\boldsymbol{k}'}\rho P_{\boldsymbol{k},\boldsymbol{k}'}Q'P_{\boldsymbol{k},\boldsymbol{k}'} \\
&= \sum_{\boldsymbol{k},\boldsymbol{k}'} Z_{\boldsymbol{k}}X_{\boldsymbol{k}'}Z_{\boldsymbol{a}}X_{\boldsymbol{a}'}Z_{\boldsymbol{k}}X_{\boldsymbol{k}'}\rho Z_{\boldsymbol{k}}X_{\boldsymbol{k}'}Z_{\boldsymbol{b}}X_{\boldsymbol{b}'}Z_{\boldsymbol{k}}X_{\boldsymbol{k}'} && (73)\\
&= \sum_{\boldsymbol{k},\boldsymbol{k}'} Z_{\boldsymbol{k}}(X_{\boldsymbol{k}'}Z_{\boldsymbol{a}}X_{\boldsymbol{k}'})X_{\boldsymbol{a}'}Z_{\boldsymbol{k}}\rho Z_{\boldsymbol{k}}(X_{\boldsymbol{k}'}Z_{\boldsymbol{b}}X_{\boldsymbol{k}'})X_{\boldsymbol{b}'}Z_{\boldsymbol{k}} && (74)\\
&= \sum_{\boldsymbol{k},\boldsymbol{k}'} Z_{\boldsymbol{k}}((-1)^{\boldsymbol{k}'\cdot\boldsymbol{a}}Z_{\boldsymbol{a}})X_{\boldsymbol{a}'}Z_{\boldsymbol{k}}\rho Z_{\boldsymbol{k}}((-1)^{\boldsymbol{k}'\cdot\boldsymbol{b}}Z_{\boldsymbol{b}})X_{\boldsymbol{b}'}Z_{\boldsymbol{k}} && (75)\\
&= \sum_{\boldsymbol{k},\boldsymbol{k}'} (-1)^{\boldsymbol{k}'\cdot(\boldsymbol{a}\oplus\boldsymbol{b})}Z_{\boldsymbol{a}}(Z_{\boldsymbol{k}}X_{\boldsymbol{a}'}Z_{\boldsymbol{k}})\rho Z_{\boldsymbol{b}}(Z_{\boldsymbol{k}}X_{\boldsymbol{b}'}Z_{\boldsymbol{k}}) && (76)\\
&= \sum_{\boldsymbol{k}'} (-1)^{\boldsymbol{k}'\cdot(\boldsymbol{a}\oplus\boldsymbol{b})} \sum_{\boldsymbol{k}} (-1)^{\boldsymbol{k}\cdot(\boldsymbol{a}'\oplus\boldsymbol{b}')}Z_{\boldsymbol{a}}X_{\boldsymbol{a}'}\rho Z_{\boldsymbol{b}}X_{\boldsymbol{b}'} && (77)
\end{aligned}$$

If either $\boldsymbol{a} \neq \boldsymbol{b}$ or $\boldsymbol{a}' \neq \boldsymbol{b}'$ the summation $\sum_{\boldsymbol{k}'}((-1)^{\boldsymbol{k}'\cdot(\boldsymbol{a}\oplus\boldsymbol{b})})$ or $\sum_{\boldsymbol{k}}((-1)^{\boldsymbol{k}\cdot(\boldsymbol{a}'\oplus\boldsymbol{b}')})$ is equal to zero respectively, because (in either case) exactly half of the elements of the summation will be -1 and half will be 1. Therefore, since our assumption was that either $\boldsymbol{a} \neq \boldsymbol{b}$ or $\boldsymbol{a}' \neq \boldsymbol{b}'$ or both, the whole expression equals zero.

□

Corollary 1.

$$\sum_{i=1}^{2^n} P_iQP_i\rho P_iQ'P_i = 0, \text{ if } Q \neq Q' \tag{78}$$

where ρ is a matrix of dimension $2^n \times 2^n$, Q, Q' are two arbitrary n-fold tensor products of Pauli X and identity operators $\{I, X\}$, and $\{P_i\}$ is the set of all n-fold tensor products of Pauli Z and identity operators $\{I, Z\}$.

Proof. Since $Q \neq Q'$, Lemma 1 gives

$$\sum_{i=1}^{4^n} P_iQP_i\rho P_iQ'P_i = 0 \tag{79}$$

where $\{P_i\}$ is the set of all n-fold tensor products of the Pauli operators and the identity $\{I, X, Y, Z\}$. But since Q and Q' have only identity and Pauli X tensor elements the Pauli X operators of $\{P_i\}$ commute with Q and Q' on each side and give identity. □